

DEX-0209



CB

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 15/12, C07K 14/47, 16/18, G01N 33/566, C12Q 1/68, C12N 15/11, 15/62, A01K 67/027, A61K 38/00</b>		<b>A2</b>	(11) International Publication Number: <b>WO 00/58473</b>
			(43) International Publication Date: 5 October 2000 (05.10.00)
(21) International Application Number: <b>PCT/US00/08621</b>		(72) Inventors; and	
(22) International Filing Date: 31 March 2000 (31.03.00)		(75) Inventors/Applicants (for US only): SHIMKETS, Richard, A. [US/US]; 191 Leete Street, West Haven, CT 06516 (US). LEACH, Martin [GB/US]; 884 School Street, Webster, MA 01570 (US).	
(30) Priority Data:		(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).	
60/127,607 31 March 1999 (31.03.99) US			
60/127,636 2 April 1999 (02.04.99) US			
60/127,728 5 April 1999 (05.04.99) US			
09/540,763 30 March 2000 (30.03.00) US			
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
US 60/127,607 (CIP)			
Filed on 31 March 1999 (31.03.99)			
US 60/127,636 (CIP)			
Filed on 2 April 1999 (02.04.99)			
US 60/127,728 (CIP)			
Filed on 5 April 1999 (05.04.99)			
US 09/540,763 (CIP)			
Filed on 30 March 2000 (30.03.00)			
(71) Applicant (for all designated States except US): CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).		Published Without international search report and to be republished upon receipt of that report.	
(54) Title: NUCLEIC ACIDS INCLUDING OPEN READING FRAMES ENCODING POLYPEPTIDES; "ORFX"			
(57) Abstract			
<p>The present invention provides open reading frames ORFX, encoding isolated polypeptides, as well as polynucleotides encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptides, polynucleotides or antibodies. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



## NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

5

## BACKGROUND OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides encoded thereby, and methods of using these nucleic acids and polypeptides.

10

## SUMMARY OF THE INVENTION

The invention is based in part on the discovery of nucleic acids that include open reading frames encoding novel polypeptides, and on the polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "ORFX".

15

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule (SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1-3161), that encodes novel polypeptide, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1-3161. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

20

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a recombinant expression vector comprising any of the nucleic acid molecules described above.

25

In another aspect, the invention includes a pharmaceutical composition that includes an ORFX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified ORF polypeptide, *e.g.*, any of the ORFX polypeptides encoded by an ORFX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a ORFX polypeptide and a pharmaceutically acceptable carrier or diluent.

5 In a still a further aspect, the invention provides an antibody that binds specifically to an ORFX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including ORFX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a  
10 polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an ORFX polypeptide by providing a cell containing a ORFX nucleic acid, *e.g.*, a vector that includes a ORFX nucleic  
15 acid, and culturing the cell under conditions sufficient to express the ORFX polypeptide encoded by the nucleic acid. The expressed ORFX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous ORFX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an ORFX polypeptide or nucleic  
20 acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a ORFX polypeptide by contacting ORFX polypeptide with a compound and determining whether the ORFX polypeptide activity is modified.

25 The invention is also directed to compounds that modulate ORFX polypeptide activity identified by contacting a ORFX polypeptide with the compound and determining whether the compound modifies activity of the ORFX polypeptide, binds to the ORFX polypeptide, or binds to a nucleic acid molecule encoding a ORFX polypeptide.

In a another aspect, the invention provides a method of determining the presence of or  
30 predisposition of an ORFX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of ORFX polypeptide in the subject sample.

The amount of ORFX polypeptide in the subject sample is then compared to the amount of ORFX polypeptide in a control sample. An alteration in the amount of ORFX polypeptide in the subject protein sample relative to the amount of ORFX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the ORFX is detected using a ORFX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the ORFX nucleic acid in the subject nucleic acid sample. The amount of ORFX nucleic acid sample in the subject nucleic acid is then compared to the amount of an ORFX nucleic acid in a control sample. An alteration in the amount of ORFX nucleic acid in the sample relative to the amount of ORFX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides method of treating or preventing or delaying a ORFX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a ORFX nucleic acid, a ORFX polypeptide, or an ORFX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

## DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polypeptides and nucleotides encoded thereby. The polynucleotides and their encoded polypeptides can be grouped according to the functions played by their gene products. Such functions include, structural proteins, proteins from which associated with metabolic pathways fatty acid metabolism, glycolysis, intermediary metabolism, calcium metabolism, proteases, and amino acid metabolism, etc.

Included in the invention are 3161 novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to as "ORFX nucleic acids" or ORFX polynucleotides" and the corresponding encoded polypeptide is referred to as a "ORFX polypeptide" or ORFX protein". For example, an ORFX nucleic acid according to the invention is a nucleic acid including an ORF1 nucleic acid, and an ORF polypeptide according to the invention is a polypeptide that includes the amino acid sequence of an ORF1 polypeptide. Unless indicated otherwise, "ORFX" is meant to refer to any of the ORF1-3161 sequences disclosed herein.

Table 1 provides a summary of the ORFX nucleic acids and their encoded polypeptides are summarized in Table 1. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention is provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 1 of Table 1, entitled "ORF #", denotes an ORF number assigned to a nucleic acid containing an open reading frame according to the invention.

Column 2 of Table 1, entitled "Internal Identification number (Nucleic Acid Sequence Identification Number, Polypeptide Sequence Identification Number), provides an internal identification number for the indicated ORF, along with sequence identification numbers (SEQ ID NOs.) corresponding to the indicated ORF. In general, for an ORF<sub>n</sub> according to the invention (wherein *n* is any integer from 1 to 3161), a nucleic acid corresponding to the ORF is SEQ ID NO:2*n*-1, and an amino acid sequence encoded by the ORF is SEQ ID NO:2*n*. For example, a nucleic acid sequence corresponding to an ORF1 nucleic acid is SEQ ID NO:1, and a polypeptide sequence corresponding to an ORF1 polypeptide is SEQ ID NO:2. Similarly, a

nucleic acid sequence corresponding to an ORF4 nucleic acid is SEQ ID NO:7, and a polypeptide sequence corresponding to an ORF4 polypeptide is SEQ ID NO:8; a nucleic acid sequence corresponding to an ORF198 nucleic acid sequence is SEQ ID NO:395, and a polypeptide sequence corresponding to an ORF198 polypeptide is SEQ ID NO:396. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention are provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 2 of Table 1, entitled "Protein Similarity", lists previously described proteins that are related to polypeptides encoded by the ORFs. Genbank identifiers for the previously described proteins are provided. These can be retrieved from <http://www.ncbi.nlm.nih.gov/>.

To determine similarity to previously described proteins, polypeptides encoded by ORFX DNA sequences were tested using the Framesearch Algorithm against a nonredundant version of the GenPept Database from NCBI/Genbank. DNA sequences that had a score of '90' or above (Framesearch algorithm score, Edelman et. al. GCG Genetics) to a known protein were selected. Open reading frames were extended beyond the region of the protein matched using standard DNA translation and codon tables. Novel proteins that lacked a protein match were translated against the standard genetic codons and proteins with an ORF at least 80 amino acids and containing a Methionine start are included in the Table.

Column 3 of Table 3, entitled "Protein Domains", lists previously described protein domains, designated by pfam entries, that are present in polypeptides encoded by the ORFs. Also included in column 3 are proteins in which these domains are present. The pfam entries can be retrieved from <http://pfam.wustl.edu/>. DNA sequences were translated in all six frames and tested using the Hmmer Algorithm against the Pfam Database (References to the algorithm and Pfam database can be found at <http://pfam.wustl.edu/>). Translated DNA sequences that matched a protein domain entry in the Pfam database AND had a score of '7.5' were selected.

Column 4 of Table 3, entitled "Protein Classification", lists the type of classification assigned for the protein, based on its homology. Examples of proteins in the classification include the following proteins:

### **Amylases**

Amylase is responsible for endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides. Variations in amylase gene may be indicative of delayed maturation and of various amylase producing neoplasms and carcinomas.

### 5           **Amyloid**

The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL). The synthesis of certain members of the family is greatly increased in inflammation. Prolonged elevation of plasma SAA levels, as in chronic inflammation, 15 results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. Amyloid selectively inhibits insulin-stimulated glucose utilization and glycogen deposition in muscle, while not affecting adipocyte glucose metabolism. Deposition of fibrillar amyloid proteins intraneuronally, as neurofibrillary tangles, extracellularly, as plaques and in blood vessels, is characteristic of both Alzheimer's disease and aged Down's syndrome. 15 Amyloid deposition is also associated with type II diabetes mellitus.

### **Angiopoeitin**

Members of the angiopoetin/fibrinogen family have been shown to stimulate the generation of new blood vessels, inhibit the generation of new blood vessels, and perform several roles in blood clotting. This generation of new blood vessels, called angiogenesis, is also an 20 essential step in tumor growth in order for the tumor to get the blood supply it needs to expand. Variation in these genes may be predictive of any form of heart disease, numerous blood clotting disorders, stroke, hypertension and predisposition to tumor formation and metastasis. In particular, these variants may be predictive of the response to various antihypertensive drugs and chemotherapeutic and anti-tumor agents.

### 25           **Apoptosis-related proteins**

Active cell suicide (apoptosis) is induced by events such as growth factor withdrawal and toxins. It is controlled by regulators, which have either an inhibitory effect on programmed cell

death (anti-apoptotic) or block the protective effect of inhibitors (pro-apoptotic). Many viruses have found a way of countering defensive apoptosis by encoding their own anti-apoptosis genes preventing their target-cells from dying too soon. Variants of apoptosis related genes may be useful in formulation of anti-aging drugs.

5           **Cadherin, Cyclin, Polymerase, Oncogenes, Histones, Kinases**

Members of the cell division/cell cycle pathways such as cyclins, many transcription factors and kinases, DNA polymerases, histones, helicases and other oncogenes play a critical role in carcinogenesis where the uncontrolled proliferation of cells leads to tumor formation and eventually metastasis. Variation in these genes may be predictive of predisposition to any form  
10 of cancer, from increased risk of tumor formation to increased rate of metastasis. In particular, these variants may be predictive of the response to various chemotherapeutic and anti-tumor agents.

**Colony-stimulating factor-related proteins**

Granulocyte/macrophage colony-stimulating factors are cytokines that act in  
15 hematopoiesis by controlling the production, differentiation, and function of 2 related white cell populations of the blood, the granulocytes and the monocytes-macrophages.

**Complement-related proteins**

Complement proteins are immune associated cytotoxic agents, acting in a chain reaction to exterminate target cells to that were opsonized (primed) with antibodies, by forming a  
20 membrane attack complex (MAC). The mechanism of killing is by opening pores in the target cell membrane. Variations in 20 complement genes or their inhibitors are associated with many autoimmune disorders. Modified serum levels of complement products cause edemas of various tissues, lupus (SLE), vasculitis, glomerulonephritis, renal failure, hemolytic anemia, thrombocytopenia, and arthritis. They interfere with mechanisms of ADCC (antibody dependent  
25 cell cytotoxicity), severely impair immune competence and reduce phagocytic ability. Variants of complement genes may also be indicative of type I diabetes mellitus, meningitis neurological disorders such as nemaline myopathy, neonatal hypotonia, muscular disorders such as congenital myopathy and other diseases.

### **Cytochrome**

The respiratory chain is a key biochemical pathway which is essential to all aerobic cells. There are five different cytochromes involved in the chain. These are heme bound proteins which serve as electron carriers. Modifications in these genes may be predictive of ataxia areflexia, dementia and myopathic and neuropathic changes in muscles. Also, association with various types of solid tumors.

### **Kinesins**

Kinesins are tubulin molecular motors that function to transport organelles within cells and to move chromosomes along microtubules during cell division. Modifications of these genes may be indicative of neurological disorders such as Pick disease of the brain, tuberous sclerosis.

### **Cytokines, Interferon, Interleukin**

Members of the cytokine families are known for their potent ability to stimulate cell growth and division even at low concentrations. Cytokines such as erythropoietin are cell-specific in their growth stimulation; erythropoietin is useful for the stimulation of the proliferation of erythroblasts. Variants in cytokines may be predictive for a wide variety of diseases, including cancer predisposition.

### **G-protein coupled receptors**

G-protein coupled receptors (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. Alterations in genes coding for G-coupled proteins may be involved in and indicative of a vast number of physiological conditions. These include blood pressure regulation, renal dysfunctions, male infertility, dopamine associated cognitive, emotional, and endocrine functions, hypercalcemia, chondrodysplasia and osteoporosis, pseudohypoparathyroidism, growth retardation and dwarfism.



**Thioesterases**

Eukaryotic thiol proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Variants of thioester associated genes may be predictive of neuronal disorders and mental illnesses such as Ceroid Lipoffiscinosis, Neuronal 1, Infantile, Santavuori disease and more.

The key to the molecule type is as follows:

10	<b>Abbrev:</b>	<b>Title:</b>
	amylase	amylase protein
	amylaseinhib	amylase inhibitor
	amyloid	amyloid protein
15	apoptosis	apoptosis associated protein
	apoptosisinhib	apoptosis inhibitors
	apoptosisrecep	apoptosis receptors
	ATPase_associated	ATPase associated protein
	biotindp	biotin dependent enzyme/protein
20	cadherin	cadherin protein
	calcium_channel	calcium channel protein
	carboxylase	carboxylase protein
	cathepsin	cathepsin/carboxypeptidases
	cathepsininhib	cathepsin/carboxypeptidase inhibitor
25	chloride_channel	chloride channel protein
	collagen	collagen
	complement	complement protein
	complementrecept	complement receptor protein
	complementinhib	complement inhibitor
30	csf	colony stimulating factor
	csfrecept	colony stimulating factor receptor
	cyclin	cyclin protein
	cyto450	cytochrome p450 protein
	cytochrome	cytochrome related protein
35	deaminase	deaminase
	dehydrogenase	dehydrogenase
	desaturase	desaturase
	dna_rna_bind	DNA/RNA binding protein/factor
	dna_rna_inhib	DNA/RNA binding protein/factor inhibitor
40	dynein	dynein

	elastase	elastase
	elastaseinhib	elastase inhibitor
	eph	EPH family of tyrosine kinases
	esterase	esterase
5	esteraseinhib	esterase inhibitor
	fgf	fibroblast growth factor
	fgfreceptor	fibroblast growth factor receptor
	gaba	GABA receptor
	glucoamylase	glucoamylase
10	glucoronidase	glucoronidase
	glycoprotein	glycoprotein
	Guanylyl	guanylylate cyclase
	helicase	helicase
	histone	histone
15	HOM	homologous
	homeobox	homeobox protein
	hydrolase	hydrolase
	hydroxysteroid	hydroxysteroid associated protein
	hypoxanthine	hypoxanthine associated protein
20	immunoglob	immunoglobulin
	immunoglobrecept	immunoglobulin receptor
	interferon	interferon
	interleukin	interleukin
	interleukinrecept	interleukin receptor
25	isomerase	isomerase
	isomeraseinhibitor	isomerase inhibitor
	isomerasereceptor	isomerase receptor
	kinase	kinase
	kinaseinhibitor	kinase inhibitor
30	kinasereceptor	kinase receptor
	kinesin	kinesin
	laminin	laminin associated protein
	lipase	lipase
	metallothionein	metallothionein
35	MHC	major histocompatibility complex
	misc_channel	miscellaneous channel
	ngf	nerve growth factor
	nuci_recpt	nuclear receptor
	nuclease	nuclease
40	oncogene	oncogene associated protein
	oxidase	oxidase
	oxygenase	oxygenase
	peptidase	peptidase
	peroxidase	peroxidase
45	phosphatase	phosphatase
	phosphataseinhib	phosphatase inhibitor

	phosphorylase	phosphorylase
	PIR	PIR DATABASE (release 56, 29-OCT-1998)
5	polymerase	polymerase
	potassium_channel	potassium channel protein
	prostaglandin	prostaglandin
	protease	protease
	proteaseinhib	protease inhibitor
10	reductase	reductase
	ribosomalprot	ribosomal associated protein
	RTR	EMBLDATABASE translated entries not to be incorporated into SWISS-PROT (20-JUL-1998)
	SIM	similar
15	SPTR	EMBL DATABASE translated entries to be incorporated into SWISS-PROT (20-JUL-1998)
	struct	structural associated protein
20	sulfotransferase	sulfotransferase
	SWP	SWISS-PROT DATABASE (release 18-OCT-1998)
	SWPN	SWISS-PROT Update (release 11-NOV-98)
	synthase	synthase
25	tgf	transforming growth factor
	tgfreceptor	transforming growth factor receptor
	thioesterase	thioesterase
	thiolase	thiolase
	tm7	seven transmembrane domain G-protein coupled receptor
30	tnf	necrosis factor receptor
	traffic	tumor necrosis factor
	tnfreceptor	tumor trafficking associated protein
	TRN	EMBL DATABASE translated entries update (20-JUL-1998)
35	transcriptfactor	transcription factor
	transferase	transferase
	transport	transport protein
	tubulin	tubulin
40	ubiquitin	ubiquitin
	unclassified	Protein not categorized into one of the aforementioned protein families
	water channel	water channel protein

Column 5 of Table 1, entitled, "Cells or Tissues in Which Gene is Expressed", denotes tissues, represented by five digit numbers, in which RNA homologous to the ORF nucleic acid sequences is present. Tissues or cells corresponding to the numbers are provided in Table 2.

5 ORFX nucleic acids, and their encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, various ORFX nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families indicated in Table 1, and/or according to the presence of domains and sequence relatedness to previously described proteins as summarized in Table 1.

10 ORFX nucleic acids and polypeptides according to the invention can also be used to identify cell types listed in Table 1 for an indicated ORFX according to the invention. Additional utilities for ORFX nucleic acids and polypeptides according to the invention are disclosed herein.

### ORFX Nucleic Acids

15 The novel nucleic acids of the invention include those that encode an ORFX or ORFX-like protein, or biologically active portions thereof. The nucleic acids include nucleic acids encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO:2*n*, wherein  $n = 1$  to 3161. The encoded polypeptides can thus include, *e.g.*, the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, . . . , 6310, 6312, 6314, 6316, 6318, 6320, and/or 6322.

20 In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO:2*n* (wherein  $n = 1$  to 3161) includes the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein  $n = 1$  to 3161), or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of any of SEQ ID NO:2*n*-1 (wherein  $n = 1$  to 3161), or a fragment thereof, any of whose bases may be changed from the  
25 disclosed sequence while still encoding a protein that maintains its ORFX-like activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein  $n = 1$  to 3161), including fragments, derivatives,

analog and homolog thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify ORFX-encoding nucleic acids (*e.g.*, ORFX mRNA) and fragments for use as  
5 polymerase chain reaction (PCR) primers for the amplification or mutation of ORFX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is  
10 double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and  
15 much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic  
20 acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism  
25 from which the nucleic acid is derived. For example, in various embodiments, the isolated ORFX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular  
30 material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n*=1 to 3161) as a hybridization probe, ORFX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to ORFX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in

is one that is sufficiently complementary to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161),, thereby forming a stable duplex.

- 5           As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect.
- 10 Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

- Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of ORFX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively,
- 20 and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but
- 25 differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

- Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or
- 30 analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the

invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of ORFX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a ORFX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human ORFX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) as well as a polypeptide having ORFX activity. Biological activities of the ORFX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human ORFX polypeptide.

The nucleotide sequence determined from the cloning of the human ORFX gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed and/or cloning ORFX homologues in other cell types, *e.g.*, from other tissues, as well as ORFX homologues from other mammals. The probe/primer typically comprises a substantially purified



oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161); or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161);  
5 or of a naturally occurring mutant of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161).

Probes based on the human ORFX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be  
10 used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a ORFX protein, such as by measuring a level of a ORFX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting ORFX mRNA levels or determining whether a genomic ORFX gene has been mutated or deleted.

"A polypeptide having a biologically active portion of ORFX" refers to polypeptides  
15 exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of ORFX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), that encodes a polypeptide having a ORFX biological activity (biological activities of the ORFX  
20 proteins are summarized in Table 1), expressing the encoded portion of ORFX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of ORFX. For example, a nucleic acid fragment encoding a biologically active portion of ORFX can optionally include a domain as shown in Table 1, column 4.

#### 25 **ORFX variants**

The invention further encompasses nucleic acid molecules that differ from the disclosed ORFX nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same ORFX protein as that encoded by the nucleotide sequence shown in SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of  
30 the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

In addition to the human ORFX nucleotide sequence shown in any of SEQ ID NO:2 $n$ -1 (wherein  $n = 1$  to 3161), it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of ORFX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the ORFX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a ORFX protein, preferably a mammalian ORFX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the ORFX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in ORFX that are the result of natural allelic variation and that do not alter the functional activity of ORFX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding ORFX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of SEQ ID NO:2 $n$ -1 (wherein  $n = 1$  to 3161), are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the ORFX cDNAs of the invention can be isolated based on their homology to the human ORFX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2 $n$ -1 (wherein  $n = 1$  to 3161). In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding ORFX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at  $T_m$ , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of any of SEQ ID NO:2 $n$ -1 (wherein  $n = 1$  to 3161) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2 $n$ -1 (wherein  $n = 1$  to 3161), or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are

hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and  
5 Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A  
10 non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at  
15 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

#### Conservative mutations

In addition to naturally-occurring allelic variants of the ORFX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), thereby leading to changes in the amino acid sequence of the encoded ORFX protein, without altering the functional ability of the ORFX protein. For example, nucleotide substitutions  
25 leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of ORFX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the ORFX proteins of the present  
30 invention, are predicted to be particularly unamenable to alteration.

Amino acid residues that are conserved among members of an ORFX family members are predicted to be less amenable to alteration. For example, an ORFX protein according to the present invention can contain at least one domain (*e.g.*, as shown in Table 1) that is a typically conserved region in an ORFX family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the ORFX family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding ORFX proteins that contain changes in amino acid residues that are not essential for activity. Such ORFX proteins differ in amino acid sequence from any of any of SEQ ID NO:2 $n$  (wherein  $n = 1$  to 3161), yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO:2 $n$  (wherein  $n = 1$  to 3161). Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to any of SEQ ID NO:2 $n$  (wherein  $n = 1$  to 3161), more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO:2.

An isolated nucleic acid molecule encoding a ORFX protein homologous to the protein of any of SEQ ID NO:2 $n$  (wherein  $n = 1$  to 3161) can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence, *i.e.* SEQ ID NO:2 $n-1$  for the corresponding  $n$ , such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO:2 $n-1$  (wherein  $n = 1$  to 3161) by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline,

phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in ORFX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a ORFX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ORFX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant ORFX protein can be assayed for (1) the ability to form protein:protein interactions with other ORFX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant ORFX protein and a ORFX receptor; (3) the ability of a mutant ORFX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind BRA protein; or (5) the ability to specifically bind an anti-ORFX protein antibody.

### Antisense

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire ORFX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a ORFX protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) or antisense nucleic acids complementary to a ORFX nucleic acid sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid

residues (*e.g.*, the protein coding region of a human ORFX that corresponds to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161)). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding ORFX disclosed herein (*e.g.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of ORFX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of ORFX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of ORFX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil,

3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a ORFX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

#### Ribozymes and PNA moieties

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are



carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme.

- 5 Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave ORFX mRNA transcripts to thereby inhibit translation of ORFX mRNA. A ribozyme having specificity for a ORFX-encoding
- 10 nucleic acid can be designed based upon the nucleotide sequence of a ORFX DNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a ORFX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742.
- 15 Alternatively, ORFX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

- Alternatively, ORFX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the ORFX (*e.g.*, the ORFX promoter and/or
- 20 enhancers) to form triple helical structures that prevent transcription of the ORFX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

- In various embodiments, the nucleic acids of ORFX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of
- 25 the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has
- 30 been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide

synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of ORFX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of ORFX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of ORFX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of ORFX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-1124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or

the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

### ORFX polypeptides

The novel protein of the invention includes the ORFX-like protein whose sequence is provided in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in FIG. 1 while still encoding a protein that maintains its ORFX-like activities and physiological functions, or a functional fragment thereof. For example, the invention includes the polypeptides encoded by the variant ORFX nucleic acids described above. In the mutant or variant protein, up to 20% or more of the residues may be so changed.

In general, an ORFX -like variant that preserves ORFX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above. Furthermore, without limiting the scope of the invention, positions of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) may be substituted such that a mutant or variant protein may include one or more substitutions

The invention also includes isolated ORFX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-ORFX antibodies. In one embodiment, native ORFX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, ORFX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a ORFX

protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the ORFX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of ORFX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of ORFX protein having less than about 30% (by dry weight) of non-ORFX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-ORFX protein, still more preferably less than about 10% of non-ORFX protein, and most preferably less than about 5% non-ORFX protein. When the ORFX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein having less than about 30% (by dry weight) of chemical precursors or non-ORFX chemicals, more preferably less than about 20% chemical precursors or non-ORFX chemicals, still more preferably less than about 10% chemical precursors or non-ORFX chemicals, and most preferably less than about 5% chemical precursors or non-ORFX chemicals.

Biologically active portions of a ORFX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the ORFX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2 that include fewer amino acids than the full length ORFX proteins, and exhibit at least one activity of a ORFX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the ORFX protein. A biologically active portion of a ORFX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a ORFX protein of the present invention may contain at least one of the above-identified domains conserved between the FGF family of proteins. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native ORFX protein.

In an embodiment, the ORFX protein has an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). In other embodiments, the ORFX protein is substantially homologous to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) and retains the functional activity of the protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the ORFX protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) and retains the functional activity of the ORFX proteins of the corresponding polypeptide having the sequence of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

#### **Determining homology between two or more sequences**

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch 1970 J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a

degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161).

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

#### **Chimeric and fusion proteins**

The invention also provides ORFX chimeric or fusion proteins. As used herein, a ORFX "chimeric protein" or "fusion protein" includes a ORFX polypeptide operatively linked to a non-ORFX polypeptide. A "ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to ORFX, whereas a "non-ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the ORFX protein, *e.g.*, a protein that is different from the ORFX protein and that is derived from the same or a different organism. Within a ORFX fusion protein the ORFX polypeptide can correspond to all or a portion of a ORFX protein. In one embodiment, a ORFX fusion protein comprises at least one biologically active portion of a ORFX protein. In another embodiment, a ORFX fusion protein comprises at least two biologically active portions of a

ORFX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the ORFX polypeptide and the non-ORFX polypeptide are fused in-frame to each other. The non-ORFX polypeptide can be fused to the N-terminus or C-terminus of the ORFX polypeptide.

5           For example, in one embodiment a ORFX fusion protein comprises a ORFX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate ORFX activity (such assays are described in detail below).

10           In another embodiment, the fusion protein is a GST-ORFX fusion protein in which the ORFX sequences are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant ORFX.

15           In yet another embodiment, the fusion protein is a ORFX protein containing a heterologous signal sequence at its N-terminus. For example, the native ORFX signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of ORFX can be increased through use of a heterologous signal sequence.

20           In another embodiment, the fusion protein is a ORFX-immunoglobulin fusion protein in which the ORFX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The ORFX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ORFX ligand and a ORFX protein on the surface of a cell, to thereby suppress ORFX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated ORFX ligand of the invention is an ORFX receptor. The ORFX-immunoglobulin fusion proteins can be used to modulate the bioavailability of a ORFX cognate ligand. Inhibition of the ORFX ligand/ORFX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g., promoting or inhibiting) cell survival. Moreover, the ORFX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-ORFX antibodies in a subject, to purify ORFX ligands, and in screening assays to identify molecules that inhibit the interaction of  
30   ORFX with a ORFX ligand.

A ORFX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A ORFX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the ORFX protein.

#### **ORFX agonists and antagonists**

The present invention also pertains to variants of the ORFX proteins that function as either ORFX agonists (mimetics) or as ORFX antagonists. Variants of the ORFX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the ORFX protein. An agonist of the ORFX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the ORFX protein. An antagonist of the ORFX protein can inhibit one or more of the activities of the naturally occurring form of the ORFX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the ORFX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the ORFX proteins.

Variants of the ORFX protein that function as either ORFX agonists (mimetics) or as ORFX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the ORFX protein for ORFX protein agonist or antagonist activity. In one



embodiment, a variegated library of ORFX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of ORFX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential ORFX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of ORFX sequences therein. There are a variety of methods which can be used to produce libraries of potential ORFX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential ORFX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

#### Polypeptide libraries

In addition, libraries of fragments of the ORFX protein coding sequence can be used to generate a variegated population of ORFX fragments for screening and subsequent selection of variants of a ORFX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a ORFX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the ORFX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of ORFX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors,

transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify ORFX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

### Anti-ORFX Antibodies

The invention further encompasses antibodies and antibody fragments, such as F<sub>ab</sub> or (F<sub>ab</sub>)<sub>2</sub>, that bind immunospecifically to any of the proteins of the invention.

An isolated ORFX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind ORFX using standard techniques for polyclonal and monoclonal antibody preparation. Full-length ORFX protein can be used. Alternatively, the invention provides antigenic peptide fragments of ORFX for use as immunogens. The antigenic peptide of ORFX comprises at least 4 amino acid residues of the amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). The antigenic peptide encompasses an epitope of ORFX such that an antibody raised against the peptide forms a specific immune complex with ORFX. The antigenic peptide may comprise at least 6 aa residues, at least 8 aa residues, at least 10 aa residues, at least 15 aa residues, at least 20 aa residues, or at least 30 aa residues. In one embodiment of the invention, the antigenic peptide comprises a polypeptide comprising at least 6 contiguous amino acids of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

In an embodiment of the invention, epitopes encompassed by the antigenic peptide are regions of ORFX that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

As disclosed herein, an ORFX protein sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and

immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen, such as ORFX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,  $F_{ab}$  and  $F_{(ab)2}$  fragments, and an  $F_{ab}$  expression library. In a specific embodiment, antibodies to human ORFX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a ORFX protein sequence of any of SEQ ID NO:2n (wherein  $n = 1$  to 3161) or derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly expressed ORFX protein or a chemically synthesized ORFX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against ORFX can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of ORFX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular ORFX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular ORFX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND

CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND  
 5 CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations are incorporated herein by reference in their entirety

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a ORFX protein (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see *e.g.*, Huse, *et al.*,  
 10 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a ORFX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See *e.g.*, U.S. Patent No. 5,225,539. Each of the above citations are incorporated herein by reference. Antibody fragments that contain the idiotypes to a ORFX protein may be produced by  
 15 techniques known in the art including, but not limited to: (i) an F<sub>(ab)</sub><sub>2</sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by reducing the disulfide bridges of an F<sub>(ab)</sub><sub>2</sub> fragment; (iii) an F<sub>ab</sub> fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F<sub>v</sub> fragments.

Additionally, recombinant anti-ORFX antibodies, such as chimeric and humanized  
 20 monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent  
 25 Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988), *J.*  
 30 *Natl Cancer Inst* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525;

Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060. Each of the above citations are incorporated herein by reference.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a ORFX protein is facilitated by generation of hybridomas that bind to the fragment of a ORFX protein possessing such a domain. Antibodies that are specific for one or more domains within a ORFX protein, *e.g.*, the domain spanning the first fifty amino-terminal residues specific to ORFX when compared to FGF-9, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-ORFX antibodies may be used in methods known within the art relating to the localization and/or quantitation of a ORFX protein (*e.g.*, for use in measuring levels of the ORFX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for ORFX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-ORFX antibody (*e.g.*, monoclonal antibody) can be used to isolate ORFX by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-ORFX antibody can facilitate the purification of natural ORFX from cells and of recombinantly produced ORFX expressed in host cells. Moreover, an anti-ORFX antibody can be used to detect ORFX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the ORFX protein. Anti-ORFX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or

phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### ORFX Recombinant Vectors and Host Cells

5 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding ORFX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can  
10 be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are  
15 replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the  
20 invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the  
25 recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation  
30 system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements

(e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., ORFX proteins, mutant forms of ORFX, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of ORFX in prokaryotic or eukaryotic cells. For example, ORFX can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

5 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons  
10 for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the ORFX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*,  
15 (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, ORFX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*,  
20 SF9 cells) include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J*  
25 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory,  
30 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.



In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to ORFX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," Reviews--Trends in Genetics, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant

host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, ORFX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding ORFX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) ORFX protein. Accordingly, the invention further provides methods for producing ORFX protein using the host cells of the invention. In one embodiment,

the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding ORFX has been introduced) in a suitable medium such that ORFX protein is produced. In another embodiment, the method further comprises isolating ORFX from the medium or the host cell.

## 5           **Transgenic animals**

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which ORFX-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous ORFX sequences have been introduced into their genome or homologous recombinant animals in which endogenous ORFX sequences have been altered. Such animals are useful for studying the function and/or activity of ORFX and for identifying and/or evaluating modulators of ORFX activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous ORFX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing ORFX-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human ORFX DNA sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human ORFX gene, such as a mouse ORFX gene, can be isolated based on hybridization to the human ORFX cDNA (described further above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of

expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the ORFX transgene to direct expression of ORFX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan 1986, In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the ORFX transgene in its genome and/or expression of ORFX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding ORFX can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a ORFX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the ORFX gene. The ORFX gene can be a human gene (*e.g.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)), but more preferably, is a non-human homologue of a human ORFX gene. For example, a mouse homologue of human ORFX gene of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) can be used to construct a homologous recombination vector suitable for altering an endogenous ORFX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous ORFX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous ORFX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous ORFX protein). In the homologous recombination vector, the altered portion of the ORFX gene is flanked at its 5' and 3' ends by additional nucleic acid of the ORFX gene to allow for homologous recombination to occur between the exogenous ORFX gene carried by the vector and an endogenous ORFX gene in an embryonic stem cell. The additional flanking ORFX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector. See *e.g.*, Thomas *et al.* (1987) *Cell* 51:503 for a description of

homologous recombination vectors. The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced ORFX gene has homologously recombined with the endogenous ORFX gene are selected (see *e.g.*, Li *et al.* (1992) *Cell* 69:915).

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. See *e.g.*, Bradley 1987, In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Curr Opin Biotechnol* 2:823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, *e.g.*, Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of

this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

### Pharmaceutical Compositions

The ORFX nucleic acid molecules, ORFX proteins, and anti-ORFX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a ORFX protein or anti-ORFX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use

as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder  
5 such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

10 For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated  
15 are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

20 The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible  
25 polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as  
30 pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.



It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by any of a number of routes, *e.g.*, as described in U.S. Patent Nos. 5,703,055. Delivery can thus also include, *e.g.*, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or stereotactic injection (see *e.g.*, Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### **Additional Uses and Methods of the Invention**

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: (a) screening assays; (b) detection assays (*e.g.*, chromosomal mapping, cell and tissue typing, forensic biology), (c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and (d) methods of treatment (*e.g.*, therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used to express ORFX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect ORFX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a ORFX gene, and to modulate ORFX activity, as described further below. In addition, the ORFX proteins can be used to screen drugs or compounds that modulate the ORFX activity or expression as well as to treat disorders characterized by insufficient or excessive production of ORFX protein, for

example proliferative or differentiative disorders, or production of ORFX protein forms that have decreased or aberrant activity compared to ORFX wild type protein. In addition, the anti-ORFX antibodies of the invention can be used to detect and isolate ORFX proteins and modulate ORFX activity.

5           This invention further pertains to novel agents identified by the above described screening assays and uses thereof for treatments as described herein.

### Screening Assays

          The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides,  
10   peptidomimetics, small molecules or other drugs) that bind to ORFX proteins or have a stimulatory or inhibitory effect on, for example, ORFX expression or ORFX activity.

          In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a ORFX protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained  
15   using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are  
20   applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des* 12:145).

          Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc Natl Acad Sci U.S.A.* 90:6909; Erb *et al.* (1994) *Proc Natl Acad Sci U.S.A.* 91:11422; Zuckermann *et al.* (1994) *J Med Chem* 37:2678; Cho *et al.* (1993)  
25   *Science* 261:1303; Carrell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2059; Carell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2061; and Gallop *et al.* (1994) *J Med Chem* 37:1233.

          Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), on chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner USP '409),  
30   plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and

Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc Natl Acad Sci U.S.A.* 87:6378-6382; Felici (1991) *J Mol Biol* 222:301-310; Ladner above.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a ORFX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the ORFX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the ORFX protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX or a biologically active portion thereof can be accomplished, for example, by determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule. As used herein, a "target molecule" is a molecule with which a ORFX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a ORFX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule

associated with the internal surface of a cell membrane or a cytoplasmic molecule. A ORFX target molecule can be a non-ORFX molecule or a ORFX protein or polypeptide of the present invention. In one embodiment, a ORFX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a membrane-bound ORFX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with ORFX.

Determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a ORFX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the ORFX protein or biologically active portion thereof. Binding of the test compound to the ORFX protein can be determined either directly or indirectly as described above. In one embodiment, the assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or

biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX can be accomplished, for example, by determining the ability of the ORFX protein to bind to a ORFX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of ORFX can be accomplished by determining the ability of the ORFX protein further modulate a ORFX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the ORFX protein to preferentially bind to or modulate the activity of a ORFX target molecule.

The cell-free assays of the present invention are amenable to use of both the soluble form or the membrane-bound form of ORFX. In the case of cell-free assays comprising the membrane-bound form of ORFX, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of ORFX is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl)dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either ORFX or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to ORFX, or interaction of ORFX with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For

example, GST-ORFX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or ORFX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of ORFX binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either ORFX or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated ORFX or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with ORFX or target molecules, but which do not interfere with binding of the ORFX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or ORFX trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the ORFX or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the ORFX or target molecule.

In another embodiment, modulators of ORFX expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of ORFX mRNA or protein in the cell is determined. The level of expression of ORFX mRNA or protein in the presence of the candidate compound is compared to the level of expression of ORFX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of ORFX expression based on this comparison. For example, when expression of ORFX mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of ORFX mRNA or protein expression. Alternatively, when expression of ORFX mRNA or protein is less (statistically significantly less) in the presence of the candidate

compound than in its absence, the candidate compound is identified as an inhibitor of ORFX mRNA or protein expression. The level of ORFX mRNA or protein expression in the cells can be determined by methods described herein for detecting ORFX mRNA or protein.

In yet another aspect of the invention, the ORFX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, *e.g.*, U.S. Pat. No. 5,283,317; Zervos *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) J Biol Chem 268:12046-12054; Bartel *et al.* (1993) Biotechniques 14:920-924; Iwabuchi *et al.* (1993) Oncogene 8:1693-1696; and Brent WO94/10300), to identify other proteins that bind to or interact with ORFX ("ORFX-binding proteins" or "ORFX-bp") and modulate ORFX activity. Such ORFX-binding proteins are also likely to be involved in the propagation of signals by the ORFX proteins as, for example, upstream or downstream elements of the ORFX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for ORFX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a ORFX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with ORFX.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

#### Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a

minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

The ORFX sequences of the present invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Pat. No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the ORFX sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The ORFX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), as described above, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences are used, a more appropriate number of primers for positive individual identification would be 500-2,000.



### **Predictive Medicine**

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

### **Use of Partial ORFX Sequences in Forensic Biology**

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen

found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NOs: \_\_ are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the ORFX sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of one or more of SEQ ID NO:2*n*-1 (where *n* = 1 to 3161), having a length of at least 20 bases, preferably at least 30 bases.

The ORFX sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue, etc. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such ORFX probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, ORFX primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

### **Predictive Medicine**

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic

acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

5           Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

10           Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

          These and other agents are described in further detail in the following sections.

#### **Diagnostic Assays**

15           Other conditions in which proliferation of cells plays a role include tumors, restenosis, psoriasis, Dupuytren's contracture, diabetic complications, Kaposi's sarcoma and rheumatoid arthritis.

          An ORFX polypeptide may be used to identify an interacting polypeptide a sample or tissue. The method comprises contacting the sample or tissue with ORFX, allowing formation of a complex between the ORFX polypeptide and the interacting polypeptide, and detecting the complex, if present.

20           The proteins of the invention may be used to stimulate production of antibodies specifically binding the proteins. Such antibodies may be used in immunodiagnostic procedures to detect the occurrence of the protein in a sample. The proteins of the invention may be used to stimulate cell growth and cell proliferation in conditions in which such growth would be favorable. An example would be to counteract toxic side effects of chemotherapeutic agents on, for example, hematopoiesis and platelet formation, linings of the gastrointestinal tract, and hair follicles. They may also be used to stimulate new cell growth in neurological disorders including, for example, Alzheimer's disease. Alternatively, antagonistic treatments may be administered in which an antibody specifically binding the ORFX-like proteins of the invention

would abrogate the specific growth-inducing effects of the proteins. Such antibodies may be useful, for example, in the treatment of proliferative disorders including various tumors and benign hyperplasias.

Polynucleotides or oligonucleotides corresponding to any one portion of the ORFX nucleic acids of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) may be used to detect DNA containing a corresponding ORF gene, or detect the expression of a corresponding ORFX gene, or ORFX-like gene. For example, an ORFX nucleic acid expressed in a particular cell or tissue, as noted in Table 2, can be used to identify the presence of that particular cell type.

An exemplary method for detecting the presence or absence of ORFX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes ORFX protein such that the presence of ORFX is detected in the biological sample. An agent for detecting ORFX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to ORFX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length ORFX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to ORFX mRNA or genomic DNA, as described above. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting ORFX protein is an antibody capable of binding to ORFX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect ORFX mRNA, protein, or genomic DNA in a biological sample *in vitro* as

well as *in vivo*. For example, *in vitro* techniques for detection of ORFX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of ORFX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of ORFX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of ORFX protein include introducing into a subject a labeled anti-ORFX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting ORFX protein, mRNA, or genomic DNA, such that the presence of ORFX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of ORFX protein, mRNA or genomic DNA in the control sample with the presence of ORFX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of ORFX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting ORFX protein or mRNA in a biological sample; means for determining the amount of ORFX in the sample; and means for comparing the amount of ORFX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect ORFX protein or nucleic acid.

#### **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity in, *e.g.*, proliferative or differentiative disorders such as hyperplasias, tumors, restenosis, psoriasis, Dupuytren's

contracture, diabetic complications, or rheumatoid arthritis, etc.; and glia-associated disorders such as cerebral lesions, diabetic neuropathies, cerebral edema, senile dementia, Alzheimer's disease, etc. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant ORFX expression or activity in which a test sample is obtained from a subject and ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant ORFX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as a proliferative disorder, differentiative disorder, glia-associated disorders, etc. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant ORFX expression or activity in which a test sample is obtained and ORFX protein or nucleic acid is detected (*e.g.*, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant ORFX expression or activity.)

The methods of the invention can also be used to detect genetic lesions in a ORFX gene, thereby determining if a subject with the lesioned gene is at risk for, or suffers from, a proliferative disorder, differentiative disorder, glia-associated disorder, etc. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a ORFX-protein, or the mis-expression of the ORFX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of (1) a deletion of one or more nucleotides from a ORFX gene; (2) an addition of one or more nucleotides to a ORFX gene; (3) a substitution of one or more nucleotides of a ORFX gene, (4) a chromosomal

rearrangement of a ORFX gene; (5) an alteration in the level of a messenger RNA transcript of a ORFX gene, (6) aberrant modification of a ORFX gene, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a ORFX gene, (8) a non-wild type level of a ORFX-protein, (9) allelic loss of a ORFX gene, and (10) inappropriate post-translational modification of a ORFX-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a ORFX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the ORFX-gene (see Abravaya *et al.* (1995) *Nucl Acids Res* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a ORFX gene under conditions such that hybridization and amplification of the ORFX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990, *Proc Natl Acad Sci USA* 87:1874-1878), transcriptional amplification system (Kwoh, *et al.*, 1989, *Proc Natl Acad Sci USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *BioTechnology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a ORFX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Pat. No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in ORFX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in ORFX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.* above. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the ORFX gene and detect mutations by comparing the sequence of the sample ORFX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (1977) *PNAS* 74:560 or Sanger (1977) *PNAS* 74:5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.*, (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publ. No. WO 94/16101; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Other methods for detecting mutations in the ORFX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA



heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type ORFX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al* (1988) *Proc Natl Acad Sci USA* 85:4397; Saleeba *et al* (1992) *Methods Enzymol* 217:286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in ORFX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a ORFX sequence, *e.g.*, a wild-type ORFX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in ORFX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl Acad Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control ORFX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting

alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA, rather than DNA, in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, *e.g.*, Keen *et al.* (1991) *Trends Genet* 7:5.

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, *e.g.*, Myers *et al* (1985) *Nature* 313:495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, *e.g.*, Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, *e.g.*, Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc Natl Acad. Sci USA* 86:6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, *e.g.*, Gasparini *et al* (1992) *Mol Cell Probes* 6:1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase

for amplification. See, *e.g.*, Barany (1991) *Proc Natl Acad Sci USA* 88:189. In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

5           The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a ORFX gene.

10           Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which ORFX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

#### **Pharmacogenomics**

15           Agents, or modulators that have a stimulatory or inhibitory effect on ORFX activity (*e.g.*, ORFX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, neurological, cancer-related or gestational disorders) associated with aberrant ORFX activity. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

25           Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*,  
30           Eichelbaum, 1996, *Clin Exp Pharmacol Physiol*, 23:983-985 and Linder, 1997, *Clin Chem*, 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic

conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a ORFX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

### Monitoring Clinical Efficacy

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied in basic drug screening and in clinical trials. For example, the effectiveness of an agent  
5 determined by a screening assay as described herein to increase ORFX gene expression, protein levels, or upregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting decreased ORFX gene expression, protein levels, or downregulated ORFX activity.

Alternatively, the effectiveness of an agent determined by a screening assay to decrease ORFX gene expression, protein levels, or downregulate ORFX activity, can be monitored in clinical  
10 trials of subjects exhibiting increased ORFX gene expression, protein levels, or upregulated ORFX activity. In such clinical trials, the expression or activity of ORFX and, preferably, other genes that have been implicated in, for example, a proliferative or neurological disorder, can be used as a "read out" or marker of the responsiveness of a particular cell.

For example, genes, including ORFX, that are modulated in cells by treatment with an  
15 agent (*e.g.*, compound, drug or small molecule) that modulates ORFX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of ORFX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by  
20 Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of ORFX or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the  
25 agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, nucleic acid, peptidomimetic, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a  
30 subject prior to administration of the agent; (*ii*) detecting the level of expression of a ORFX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more

post-administration samples from the subject; (iv) detecting the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the pre-administration sample with the ORFX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of ORFX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of ORFX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

#### Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant ORFX expression or activity.

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, (i) a ORFX polypeptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to a ORFX peptide; (iii) nucleic acids encoding a ORFX peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to a ORFX peptide) that are utilized to "knockout" endogenous function of a ORFX peptide by homologous recombination (see, *e.g.*, Capecchi, 1989, *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between a ORFX peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized

include, but are not limited to, a ORFX peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of a ORFX peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, etc.).

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant ORFX expression or activity, by administering to the subject an agent that modulates ORFX expression or at least one ORFX activity. Subjects at risk for a disease that is caused or contributed to by aberrant ORFX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the ORFX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of ORFX aberrancy, for example, a ORFX agonist or ORFX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Another aspect of the invention pertains to methods of modulating ORFX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of ORFX protein activity associated with the cell. An agent that modulates ORFX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a ORFX protein, a peptide, a ORFX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more ORFX protein activity. Examples of such stimulatory agents include active ORFX protein and a nucleic acid molecule encoding ORFX that has been introduced into the cell. In another embodiment, the agent inhibits one or more ORFX protein activity. Examples of such inhibitory agents include antisense ORFX nucleic acid molecules and anti-ORFX antibodies. These modulatory methods can be performed *in vitro*

(*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a ORFX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) ORFX expression or activity. In another embodiment, the method involves administering a ORFX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant ORFX expression or activity.

#### **Determination of the Biological Effect of a Therapeutic**

In various embodiments of the present invention, suitable *in vitro* or *in vivo* assays are utilized to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

#### **Malignancies**

Some ORFX polypeptides are expressed in cancerous cells (*see, e.g.*, Tables 1 and 2). Accordingly, the corresponding ORF protein is involved in the regulation of cell proliferation. Accordingly, Therapeutics of the present invention may be useful in the therapeutic or prophylactic treatment of diseases or disorders that are associated with cell hyperproliferation and/or loss of control of cell proliferation (*e.g.*, cancers, malignancies and tumors). For a review of such hyperproliferation disorders, *see e.g.*, Fishman, *et al.*, 1985. MEDICINE, 2nd ed., J.B. Lippincott Co., Philadelphia, PA.

Therapeutics of the present invention may be assayed by any method known within the art for efficacy in treating or preventing malignancies and related disorders. Such assays include,



but are not limited to, *in vitro* assays utilizing transformed cells or cells derived from the patient's tumor, as well as *in vivo* assays using animal models of cancer or malignancies. Potentially effective Therapeutics are those that, for example, inhibit the proliferation of tumor-derived or transformed cells in culture or cause a regression of tumors in animal models, in comparison to the controls.

In the practice of the present invention, once a malignancy or cancer has been shown to be amenable to treatment by modulating (*i.e.*, inhibiting, antagonizing or agonizing) activity, that cancer or malignancy may subsequently be treated or prevented by the administration of a Therapeutic that serves to modulate protein function.

#### **Premalignant conditions**

The Therapeutics of the present invention that are effective in the therapeutic or prophylactic treatment of cancer or malignancies may also be administered for the treatment of pre-malignant conditions and/or to prevent the progression of a pre-malignancy to a neoplastic or malignant state. Such prophylactic or therapeutic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia or, most particularly, dysplasia has occurred. For a review of such abnormal cell growth see *e.g.*, Robbins & Angell, 1976. BASIC PATHOLOGY, 2nd ed., W.B. Saunders Co., Philadelphia, PA.

Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in its structure or function. For example, it has been demonstrated that endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of mature or fully differentiated cell substitutes for another type of mature cell. Metaplasia may occur in epithelial or connective tissue cells. Dysplasia is generally considered a precursor of cancer, and is found mainly in the epithelia. Dysplasia is the most disorderly form of non-neoplastic cell growth, and involves a loss in individual cell uniformity and in the architectural orientation of cells. Dysplasia characteristically occurs where there exists chronic irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.

Alternatively, or in addition to the presence of abnormal cell growth characterized as hyperplasia, metaplasia, or dysplasia, the presence of one or more characteristics of a transformed or malignant phenotype displayed either *in vivo* or *in vitro* within a cell sample derived from a patient, is indicative of the desirability of prophylactic/therapeutic administration of a Therapeutic that possesses the ability to modulate activity of An aforementioned protein. Characteristics of a transformed phenotype include, but are not limited to: (i) morphological changes; (ii) looser substratum attachment; (iii) loss of cell-to-cell contact inhibition; (iv) loss of anchorage dependence; (v) protease release; (vi) increased sugar transport; (vii) decreased serum requirement; (viii) expression of fetal antigens, (ix) disappearance of the 250 kDal cell-surface protein, and the like. See *e.g.*, Richards, *et al.*, 1986. MOLECULAR PATHOLOGY, W.B. Saunders Co., Philadelphia, PA.

In a specific embodiment of the present invention, a patient that exhibits one or more of the following predisposing factors for malignancy is treated by administration of an effective amount of a Therapeutic: (i) a chromosomal translocation associated with a malignancy (*e.g.*, the Philadelphia chromosome (*bcr/abl*) for chronic myelogenous leukemia and t(14;18) for follicular lymphoma, etc.); (ii) familial polyposis or Gardner's syndrome (possible forerunners of colon cancer); (iii) monoclonal gammopathy of undetermined significance (a possible precursor of multiple myeloma) and (iv) a first degree kinship with persons having a cancer or pre-cancerous disease showing a Mendelian (genetic) inheritance pattern (*e.g.*, familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, medullary thyroid carcinoma with amyloid production and pheochromocytoma, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia and Bloom's syndrome).

In another embodiment, a Therapeutic of the present invention is administered to a human patient to prevent the progression to breast, colon, lung, pancreatic, or uterine cancer, or melanoma or sarcoma.

### **Hyperproliferative and dysproliferative disorders**

In one embodiment of the present invention, a Therapeutic is administered in the therapeutic or prophylactic treatment of hyperproliferative or benign dysproliferative disorders. The efficacy in treating or preventing hyperproliferative diseases or disorders of a Therapeutic of the present invention may be assayed by any method known within the art. Such assays include *in vitro* cell proliferation assays, *in vitro* or *in vivo* assays using animal models of hyperproliferative diseases or disorders, or the like. Potentially effective Therapeutics may, for example, promote cell proliferation in culture or cause growth or cell proliferation in animal models in comparison to controls.

Specific embodiments of the present invention are directed to the treatment or prevention of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes); treatment of keloid (hypertrophic scar) formation causing disfiguring of the skin in which the scarring process interferes with normal renewal; psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination); benign tumors; fibrocystic conditions and tissue hypertrophy (*e.g.*, benign prostatic hypertrophy).

### **Neurodegenerative disorders**

Some ORFX proteins are found in cell types have been implicated in the deregulation of cellular maturation and apoptosis, which are both characteristic of neurodegenerative disease. Accordingly, Therapeutics of the invention, particularly but not limited to those that modulate (or supply) activity of an aforementioned protein, may be effective in treating or preventing neurodegenerative disease. Therapeutics of the present invention that modulate the activity of an aforementioned protein involved in neurodegenerative disorders can be assayed by any method known in the art for efficacy in treating or preventing such neurodegenerative diseases and disorders. Such assays include *in vitro* assays for regulated cell maturation or inhibition of apoptosis or *in vivo* assays using animal models of neurodegenerative diseases or disorders, or any of the assays described below. Potentially effective Therapeutics, for example but not by way of limitation, promote regulated cell maturation and prevent cell apoptosis in culture, or reduce neurodegeneration in animal models in comparison to controls.

Once a neurodegenerative disease or disorder has been shown to be amenable to treatment by modulation activity, that neurodegenerative disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity. Such diseases include all degenerative disorders involved with aging, especially osteoarthritis and neurodegenerative disorders.

### Disorders related to organ transplantation

Some ORFX can be associated with disorders related to organ transplantation, in particular but not limited to organ rejection. Therapeutics of the invention, particularly those that modulate (or supply) activity, may be effective in treating or preventing diseases or disorders related to organ transplantation. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity of an aforementioned protein) can be assayed by any method known in the art for efficacy in treating or preventing such diseases and disorders related to organ transplantation. Such assays include *in vitro* assays for using cell culture models as described below, or *in vivo* assays using animal models of diseases and disorders related to organ transplantation, see *e.g.*, below. Potentially effective Therapeutics, for example but not by way of limitation, reduce immune rejection responses in animal models in comparison to controls.

Accordingly, once diseases and disorders related to organ transplantation are shown to be amenable to treatment by modulation of activity, such diseases or disorders can be treated or prevented by administration of a Therapeutic that modulates activity.

### Cardiovascular Disease

GENX has been implicated in cardiovascular disorders, including in atherosclerotic plaque formation. Diseases such as cardiovascular disease, including cerebral thrombosis or hemorrhage, ischemic heart or renal disease, peripheral vascular disease, or thrombosis of other major vessel, and other diseases, including diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage disease, systemic lupus erythematosus, homocysteinemia, and familial protein or lipid processing diseases, and the like, are either directly or indirectly associated with atherosclerosis. Accordingly, Therapeutics of the invention, particularly those that modulate (or supply) activity or formation may be effective in treating or preventing

atherosclerosis-associated diseases or disorders. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity) can be assayed by any method known in the art, including those described below, for efficacy in treating or preventing such diseases and disorders.

5           A vast array of animal and cell culture models exist for processes involved in atherosclerosis. A limited and non-exclusive list of animal models includes knockout mice for premature atherosclerosis (Kurabayashi and Yazaki, 1996, *Int. Angiol.* 15: 187-194), transgenic mouse models of atherosclerosis (Kappel *et al.*, 1994, *FASEB J.* 8: 583-592), antisense oligonucleotide treatment of animal models (Callow, 1995, *Curr. Opin. Cardiol.* 10: 569-576),  
10 transgenic rabbit models for atherosclerosis (Taylor, 1997, *Ann. N.Y. Acad. Sci.* 811: 146-152), hypercholesterolemic animal models (Rosenfeld, 1996, *Diabetes Res. Clin. Pract.* 30 Suppl.: 1-11), hyperlipidemic mice (Paigen *et al.*, 1994, *Curr. Opin. Lipidol.* 5: 258-264), and inhibition of lipoxygenase in animals (Sigal *et al.*, 1994, *Ann. N.Y. Acad. Sci.* 714: 211-224). In addition, *in vitro* cell models include but are not limited to monocytes exposed to low density lipoprotein  
15 (Frostegard *et al.*, 1996, *Atherosclerosis* 121: 93-103), cloned vascular smooth muscle cells (Suttles *et al.*, 1995, *Exp. Cell Res.* 218: 331-338), endothelial cell-derived chemoattractant exposed T cells (Katz *et al.*, 1994, *J. Leukoc. Biol.* 55: 567-573), cultured human aortic endothelial cells (Farber *et al.*, 1992, *Am. J. Physiol.* 262: H1088-1085), and foam cell cultures (Libby *et al.*, 1996, *Curr Opin Lipidol* 7: 330-335). Potentially effective Therapeutics, for  
20 example but not by way of limitation, reduce foam cell formation in cell culture models, or reduce atherosclerotic plaque formation in hypercholesterolemic mouse models of atherosclerosis in comparison to controls.

          Accordingly, once an atherosclerosis-associated disease or disorder has been shown to be amenable to treatment by modulation of activity or formation, that disease or disorder can be  
25 treated or prevented by administration of a Therapeutic that modulates activity.

#### **Cytokine and Cell Proliferation/Differentiation Activity**

A GENX protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered

to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+ ), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods: Assays for T-cell or thymocyte proliferation include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 3 and Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bertagnoli *et al.*, *J Immunol* 145:1706-1712, 1990; Bertagnoli *et al.*, *Cell Immunol* 133:327-341, 1991; Bertagnoli, *et al.*, *J Immunol* 149:3778-3783, 1992; Bowman *et al.*, *J Immunol* 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described by Kruisbeek and Shevach, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1, pp. 3.12.1-14, John Wiley and Sons, Toronto 1994; and by Schreiber, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan eds. Vol 1 pp. 6.8.1-8, John Wiley and Sons, Toronto 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described by Bottomly *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto 1991; deVries *et al.*, *J Exp Med* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 336:690-692, 1988; Greenberger *et al.*, *Proc Natl Acad Sci U.S.A.* 80:2931-2938, 1983; Nordan, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.6.1-5, John Wiley and Sons, Toronto 1991; Smith *et al.*, *Proc Natl Acad Sci U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, *et al.* In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto 1991; Ciarletta, *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and

cytokine production) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds., Greene Publishing Associates and Wiley-Interscience (Chapter 3Chapter 6, Chapter 7); Weinberger *et al.*, *Proc Natl Acad Sci USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur J Immun* 11:405-411, 1981; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988.

### Immune Stimulating or Suppressing Activity

A GENX protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), *e.g.*, in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by vital (*e.g.*, HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by vital, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania species., malaria species. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response.

The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or energy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon re-exposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7), *e.g.*, preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (*e.g.*, B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to energize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the



immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc Natl Acad Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and auto-antibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of auto-antibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the

patient. Another method of enhancing anti-vital immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells *in vivo*.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (*e.g.*, sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (*e.g.*, a cytoplasmic-domain truncated portion) of an MHC class I  $\alpha$  chain protein and  $\beta_2$  microglobulin protein or an MHC class II  $\alpha$  chain protein and an MHC class II  $\beta$  chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (*e.g.*, B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor

specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods: Suitable assays for thymocyte or splenocyte cytotoxicity include, without  
5 limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981; Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981;  
10 Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bowman *et al.*, *J Virology* 61:1992-1998; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Brown *et al.*, *J Immunol* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which  
15 will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J Immunol* 144:3028-3033, 1990; and Mond and Brunswick In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, (eds.) Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins  
20 that generate predominantly Th1 and CTL responses) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *J Immunol* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by  
25 dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery *et al.*, *J Immunol* 134:536-544, 1995; Inaba *et al.*, *J Exp Med* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079, 1995; Porgador *et al.*, *J Exp Med* 182:255-260, 1995; Nair *et al.*, *J Virol* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J Exp Med* 169:1255-1264, 1989; Bhardwaj *et al.*, *J Clin Investig* 94:797-807, 1994; and Inaba *et al.*, *J Exp Med* 172:631-640, 1990.  
30

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J Immunol* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Internat J Oncol* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell Immunol* 155: 111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc Nat Acad Sci USA* 88:7548-7551, 1991.

#### **Hematopoiesis Regulating Activity**

A GENX protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, *e.g.* in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (*i.e.*, in conjunction with bone marrow

transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

5           Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson *et al.* *Cellular Biology* 15:141-151, 1995; Keller *et al.*, *Mol. Cell.*  
10 *Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama *et al.*,  
15 *Proc Natl Acad Sci USA* 89:5907-5911, 1992; McNiece and Briddeli, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben *et al.*, *Exp Hematol* 22:353-359, 1994; Ploemacher, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Spoonceret *et al.*, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.*, (eds.) Vol pp.  
20 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Sutherland, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.*, (eds.) Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

### **Tissue Growth Activity**

A GENX protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for  
25 wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation  
30 employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation

induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.* for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein

may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, EPIDERMAL WOUND HEALING, pp. 71-112 (Maibach and Rovee, eds.), Year Book Medical

Publishers, Inc., Chicago, as modified by Eaglstein and Menz, *J. Invest. Dermatol* 71:382-84 (1978).

#### **Activin/Inhibin Activity**

A GENX protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-b group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc Natl Acad Sci USA* 83:3091-3095, 1986.

#### **Chemotactic/Chemokinetic Activity**

A protein of the present invention may have chemotactic or chemokinetic activity (*e.g.*, act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example,



attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population.

5 Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by following methods:

10 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Coligan *et al.*,  
15 eds. (Chapter 6.12, MEASUREMENT OF ALPHA AND BETA CHEMOKINES 6.12.1-6.12.28); Taub *et al. J Clin Invest* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Muller *et al., Eur J Immunol* 25: 1744-1748; Gruber *et al. J Immunol* 152:5860-5867, 1994; Johnston *et al., J Immunol* 153: 1762-1768, 1994.

#### **Hemostatic and Thrombolytic Activity**

20 A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for  
25 treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (*e.g.*, stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those  
30 described in: Linet *et al., J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al., Thrombosis Res.*

45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

### **Receptor/Ligand Activity**

A protein of the present invention may also demonstrate activity as receptors, receptor  
5 ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and  
ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and  
their ligands, receptor phosphatases and their ligands, receptors involved in cell—cell  
interactions and their ligands (including without limitation, cellular adhesion molecules (such as  
selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation,  
10 antigen recognition and development of cellular and humoral immune responses). Receptors and  
ligands are also useful for screening of potential peptide or small molecule inhibitors of the  
relevant receptor/ligand interaction. A protein of the present invention (including, without  
limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of  
receptor/ligand interactions.

15 The activity of a protein of the invention may, among other means, be measured by the  
following methods:

Suitable assays for receptor-ligand activity include without limitation those described in:  
CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan, *et al.*, Greene Publishing Associates and  
Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions  
20 7.28.1-7.28.22), Takai *et al.*, *Proc Natl Acad Sci USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp.*  
*Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 1989; Stoltenborg *et al.*,  
*J Immunol Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995.

### **Anti-Inflammatory Activity**

Proteins of the present invention may also exhibit anti-inflammatory activity. The  
25 anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the  
inflammatory response, by inhibiting or promoting cell—cell interactions (such as, for example,  
cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory  
process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production  
of other factors which more directly inhibit or promote an inflammatory response. Proteins  
30 exhibiting such activities can be used to treat inflammatory conditions including chronic or acute

conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

### **Tumor Inhibition Activity**

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

### **Other Activities**

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting

deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Neural disorders in general include Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumors of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury, and other neuropathies, epilepsy, and/or tremors.

## EQUIVALENTS

From the foregoing detailed description of the specific embodiments of the invention, it should be apparent that particular novel compositions and methods involving nucleic acids, polypeptides, antibodies, detection and treatment have been described. Although these particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims that follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made as a matter of routine for a person of ordinary skill in the art to the invention without departing from the spirit and scope of the invention as defined by the claims. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Table 1

ORF#	Internal Identification Number	Protein similarity	Protein domain	Protein Classification	Calls or Tissues in which Gene is Expressed
1	13076366 (1, 2)	Novel Protein sim. GBank gij4691395jemb[CAB41562.1] - (AL049727) putative large secreted protein [Streptomyces coelicolor]		UNCLASSIFIED	264636
2	80248091 (3, 4)	Novel Protein sim. GBank gij2829508ispjP71559[SUCC_MYCTU - SUCCINYL-COA SYNTHETASE BETA CHAIN (SCS-BETA)]	Contains protein domain (PF00549) - CoA-ligases	UNCLASSIFIED	264907, 264600, 264602, 264762, 264769, 264689, 264638, 264567
3	80415924 (5, 6)			UNCLASSIFIED	264910, 264604, 264634, 264905, 264636, 264691, 264907, 264692, 264629
4	82018837 (7, 8)			UNCLASSIFIED	264908, 264909, 264760, 264628, 264635
5	79970035 (9, 10)			UNCLASSIFIED	22279002, 264563
6	79842462 (11, 12)		Contains protein domain (PF00127) - Copper binding proteins, plastocyanin/azurin family	UNCLASSIFIED	264908
7	85515576 (13, 14)	Novel Protein sim. GBank gij4415929jgbjAAD20157] - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	20281099, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265006, 264512, 265009, 264910, 264595, 264596, 264758, 264603, 264604, 264760, 264762, 264683, 264766, 264767, 264689, 35695917, 264690, 264692, 264693, 33657109, 264628, 264629, 35688423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 18108385, 264563, 264564, 264566, 264486
8	56924278 (15, 16)	Novel Protein sim. GBank gij585562jspjQ06458jNIRB_KLEPN - NITRITE REDUCTASE (NAD(P)H) LARGE SUBUNIT		reductase	264907
9	79394457 (17, 18)			UNCLASSIFIED	265007, 265019, 263972
10	79556459 (19, 20)			UNCLASSIFIED	264906
11	20414027 (21, 22)			UNCLASSIFIED	264605
12	94141210 (23, 24)	Novel Protein sim. GBank gij3878145jemb[CAA9987.1] - (Z75543) similar to potassium channel protein [Caenorhabditis elegans]		misc channel	264259, 265007, 83373044
13	20750551 (25, 26)			UNCLASSIFIED	264556, 264557, 264564
14	95105114 (27, 28)	Novel Protein sim. GBank gij2832781jemb[CAA12645] - (AJ225805) inward potassium channel alpha subunit [Egeria densa]	Contains protein domain (PF00023) - Ank repeat	potassium_channel	35696286, 35696052, 264510, 35695917, 264691, 264628, 35696423, 264555, 264558, 264559, 83373044
15	20456307 (29, 30)	Novel Protein sim. GBank gij1710791jspjQ10234jRT05_SCHPO - PROBABLE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S5	Contains protein domain (PF00333) - Ribosomal protein S5	ribosomalprot	264604
16	20760356 (31, 32)				264555

17	20292744 (33, 34)	Novel Protein sim. GBank gi 1174884 sp P44391 URE1_HAEIN - UREASE ALPHA SUBUNIT (UREA AMIDOHYDROLASE)	(Contains protein domain (PF00449) - Urease	264600	
18	80246804 (35, 36)	Novel Protein sim. GBank gi 2281102 (AC002333) - SF16 isolog [Arabidopsis thaliana]		29331827, 264555, 264557, 264638, 264558	
19	80076624 (37, 38)		UNCLASSIFIED	22278996, 264907, 264910, 264600, 264693	
20	20724558 (39, 40)	Novel Protein sim. GBank gi 2508112 sp P43672 UUP_ECOLI - ABC TRANSPORTER ATP-BINDING PROTEIN UUP	transport	264602	
21	80417554 (41, 42)	Novel Protein sim. GBank gi 1730203 sp P50442 GATM_RAT - GLYCINE AMIDINOTRANSFERASE PRECURSOR (L- ARGININE:GLYCINE AMIDINOTRANSFERASE) (TRANSAMIDINASE) (AT)	UNCLASSIFIED	22278995, 264906, 265008, 265010, 265011, 264602, 264605, 264766, 264688, 21908764, 264691, 18108376, 264636, 18108387, 264486	
22	11705858 (43, 44)			264685	
23	80419178 (45, 46)	Novel Protein sim. GBank gi 1877329 emb CAB07077  - (Z92771) fadE25 [Mycobacterium tuberculosis]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	264488, 264907, 264909, 264600, 264602, 264603, 264605, 264682, 264766, 32833986, 264636, 264486	
24	20291697 (47, 48)			264600	
25	80253774 (49, 50)			264593	
26	80255394 (51, 52)		UNCLASSIFIED	22278996, 56182435, 265018, 264566	
27	80235795 (53, 54)	Novel Protein sim. GBank gi 480369 emb CAB42783.1  - (AL049841) putative 30S ribosomal protein S14 [Streptomyces coelicolor]	Contains protein domain (PF00253) - Ribosomal protein S14p/S29e	18108370, 35698423, 264635, 264555	
28	79483561 (55, 56)		UNCLASSIFIED	264638	
29	82448765 (57, 58)	Novel Protein sim. GBank gi 3122290 sp O08333 K6PF_STRCO - 6- PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE) (PHOSPHOHEXOKINASE) (ATP-PFK)	Contains protein domain (PF00355) - Phosphofructokinase	264601, 264762, 264766, 264769, 264636	
30	79199333 (59, 60)		UNCLASSIFIED	264908, 265019, 264687, 21906764, 21906766	
31	19848158 (61, 62)		UNCLASSIFIED	264534	
32	82449495 (63, 64)	Novel Protein sim. GBank gi 3560504 (AF027770) - unknown [Mycobacterium smegmatis]	UNCLASSIFIED	264905, 264605, 264762, 264766, 264687, 264689	
33	79582628 (65, 66)	Novel Protein sim. GBank gi 2129003 pir G64507 - hypothetical protein MJ1665 - Methanococcus jannaschii	UNCLASSIFIED	264687	
34	87467657 (67, 68)		UNCLASSIFIED	60432289, 264600, 264602, 264760, 18108357, 264768, 265020, 264691	
35	95005170 (69, 70)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264600, 264687, 264558, 264639	
36	18642042 (71, 72)	Novel Protein sim. GBank gi 3287739 sp P73538 BIOB_SYNY3 - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)	synthase	264566	
37	20368215 (73, 74)	Novel Protein sim. GBank gi 213134 gb AAD07126.1  - (AE000527) delta-1-pyrroline-5-carboxylate dehydrogenase [Helicobacter pylori 26695]	dehydrogenase	264603	

38	20466334 (75, 76)	Novel Protein sim. GBank gjl3805970[emb]CAA06231] - (AJ004933) periplasmic nitrate reductase, large subunit [Rhodospirillum rubrum sp.]			264605	
39	94300715 (77, 78)	Novel Protein sim. GBank gjl1929449 (L63543) - endodermin [Xenopus laevis]		Contains protein domain (PF00207) - Alpha-2-macroglobulin family	264905, 264906, 264907, 66712502, 264908, 264909, 264511, 265009, 264910, 55812038, 264758, 265011, 264762, 264682, 264763, 264764, 264766, 265022, 264693, 264628, 264631, 264634, 264635, 264555, 264636, 18108381, 264558, 18108385, 264482	
40	20635625 (78, 80)				UNCLASSIFIED	
41	80023287 (81, 82)	Novel Protein sim. GBank gjl854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 6]			264591, 35695917	
42	20724566 (83, 84)				264602	
43	20467069 (85, 86)	Novel Protein sim. GBank gjl3820584 (AF086781) - carbamoylphosphate synthetase large subunit [Zymomonas mobilis]			264605	
44	13085297 (87, 88)	Novel Protein sim. GBank gjl2494764[sp]Q50729[GUAA_MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)]		Contains protein domain (PF00958) - GMP synthase C terminal domain	264769, 264636	
45	39384711 (89, 90)	Novel Protein sim. GBank gjl1881738 (U89688) - myosin-I binding protein Acan125 [Acanthamoeba castellanii]			UNCLASSIFIED	264769, 264510, 264508
46	95003398 (91, 92)				ngf	264566
47	11696624 (93, 94)				UNCLASSIFIED	264689
48	79407218 (95, 96)				18108385, 264635, 264828	
49	21659844 (97, 98)				UNCLASSIFIED	264603
50	80503996 (99, 100)					264508, 264603, 264769, 264689, 264636, 264558, 264486
51	80255569 (101, 102)	Novel Protein sim. GBank gjl3411177 (AF076240) - MocC [Rhizobium leguminosarum bv. viciae]			UNCLASSIFIED	264593, 18108387
52	79208528 (103, 104)	Novel Protein sim. GBank gjl3914992[sp]Q26264[SM41 - HEMPU - 41 KD SPICULE MATRIX PROTEIN PRECURSOR (HSM41) (HPSMC)]			struct	264634
53	36998970 (105, 106)	Novel Protein sim. GBank gjl3980411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]			UNCLASSIFIED	264762
54	79570897 (107, 108)				UNCLASSIFIED	
55	80202703 (109, 110)	Novel Protein sim. GBank gjl1633572 (U52064) - Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]				264630, 264909, 264766
56	8758408 (111, 112)	Novel Protein sim. GBank gjl4321580[gb]AAD15785] - (AF050114) alginate lyase [Pseudomonas sp. W7]			29331824, 264102, 265018, 18108376	
57	11223388 (113, 114)			Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	264604	
					UNCLASSIFIED	264557

58	91227508 (115, 116)	Novel Protein sim. GBank gi 5616074 gb AAD45818.1 AF08194 - (AF081943) protate- derived STE20-like kinase PSK [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	56182575, 264259, 60432049, 35696052, 66712502, 264909, 265008, 265010, 265011, 264681, 29148784, 35695917, 60170615, 264691, 264692, 264693, 18108374, 35696423, 56182323, 60432113 264600, 264689, 264638
59	80077371 (117, 118)	Novel Protein sim. GBank gi 1172920 sp P45830 PFE_MYCLE - PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N- ACETYLGLUCOSAMINYLTRANSFERASE	Contains protein domain (PF00953) - Glycosyl transferase	transferase	
60	12958341 (119, 120)				264689
61	80426806 (121, 122)	Novel Protein sim. GBank gi 1710216 (U79260) - unknown [Homo sapiens]		glycoprotein	264766
62	13504966 (123, 124)				264630
63	16474553 (125, 126)			UNCLASSIFIED	265019
64	20724578 (127, 128)	Novel Protein sim. GBank gi 420945 pir A47041 - transposase homolog (Insertion element ISAE1) - Alcaligenes eutrophus		UNCLASSIFIED	264602
65	79326308 (129, 130)	Novel Protein sim. GBank gi 3122312 sp O6134 KPYK_MYCTU - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	264563
66	46854384 (131, 132)	Novel Protein sim. GBank gi 3928723 emb CAA22218  - (AL034355) putative ABC transporter [Streptomyces coelicolor]		transport	22278996, 264558
67	78952543 (133, 134)	Novel Protein sim. GBank gi 231985 sp P30234 DHA_MYCTU - ALANINE DEHYDROGENASE (40 KD ANTIGEN)		dehydrogenase	265021
68	79817382 (135, 136)				264909
69	79841764 (137, 138)			UNCLASSIFIED	264908
70	79871329 (139, 140)				264906, 264908
71	65897458 (141, 142)			UNCLASSIFIED	264602, 265021
72	87734977 (143, 144)	Novel Protein sim. GBank gi 4415926 gb AAD20157  - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 264905, 264906, 264907, 264908, 264511, 265008, 264910, 264758, 87168474, 264682, 264766, 264686, 264689, 35695917, 265021, 60170615, 264691, 33657023, 264692, 264693, 264629, 264631, 264639, 22279000
73	80025241 (145, 146)			UNCLASSIFIED	60424179, 264508, 264908, 265007, 264603, 264687, 264689, 264692, 18108387
74	20377410 (147, 148)			UNCLASSIFIED	264605
75	11819032 (149, 150)	Novel Protein sim. GBank gi 2853098 emb CAA16914  - (AL021767) vacuolar protein sorting [Schizosaccharomyces pombe]		UNCLASSIFIED	264689
76	95105303 (151, 152)	Novel Protein sim. GBank gi 4468811 emb CAB38212  - (AL035601) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	83373044, 264906, 264557
77	10144718 (153, 154)	Novel Protein sim. GBank gi 854065 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264563
78	8758258 (155, 156)			UNCLASSIFIED	264604



78	94140180 (157, 158)	Novel Protein sim. GBank gjl5689453dbj[BAA83010.1] - (AB028981) KAA1058 protein [Homo sapiens]	Contains protein domain (PF00169) - PH domain		35690286, 22278998, 29331822, 29331824, 29331825, 29331827, 264905, 264906, 264907, 66712502, 264908, 264909, 265008, 265009, 264910, 60170831, 55812038, 33109954, 265017, 265018, 264288, 264768, 56181562, 21906765, 21906769, 29148784, 265020, 264690, 264691, 264692, 264693, 60431528, 35696423, 264631, 264632, 264634, 264636, 264639, 83373044, 264564, 264586, 264587
80	82314840 (159, 160)			UNCLASSIFIED	264769, 264601, 265006, 264910, 264604, 264605, 264634, 264635, 264905, 264762, 264637, 264592, 264628, 264907, 264691, 264908, 264587, 264909, 264766
81	20467247 (161, 162)	Novel Protein sim. GBank gjl1723442isp[Q10258]YDZA, SCHPO - HYPOTHETICAL 69.0 KD PROTEIN C56F8.10 IN CHROMOSOME 1		reductase	264605
82	16331388 (163, 164)	Novel Protein sim. GBank gjl2895866 (AF045770) - methylmalonate semi-aldehyde dehydrogenase [Oryza sativa]		dehydrogenase	264587
83	94741180 (165, 166)	Novel Protein sim. GBank gjl3402673 (AC004697) - unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 264508, 264509, 264905, 264908, 264909, 264511, 264591, 264593, 264594, 264595, 264596, 264758, 264603, 264760, 264681, 18108351, 264762, 264682, 264764, 264684, 264766, 264586, 264632, 264637, 264557, 264638, 264639, 18108385, 264566
84	80355375 (167, 168)	Novel Protein sim. GBank gjl1173364isp[P45380]SAT1_RAT - SULFATE ANION TRANSPORTER 1 (CANALICULAR SULFATE TRANSPORTER) [SULFATE/CARBONATE ANTIPORTER]		transport	264508, 264906, 264907, 264908, 264909, 264910, 264760, 264763, 264764, 264766, 264768, 264769, 35695855, 264636, 264637
85	80499600 (169, 170)	Novel Protein sim. GBank gjl2120998ipr[IS70682] - glycosyltransferase homolog - Bordetella pertussis		transferase	264605, 264762, 264687, 264769, 18108374, 264636, 264486
86	39559043 (171, 172)	Novel Protein sim. GBank gjl3256023femb[CAA17228.1] - (AL021897) hypothetical protein Rv1112 [Mycobacterium tuberculosis]			264910
87	13856808 (173, 174)			UNCLASSIFIED	264093

88	95344718 (175, 176)	Novel Protein sim. GBank gjl559703[dbj BAA07552] - (D38549) ha1025 is new [Homo sapiens]		52644507, 52646365, 18108398, 65274572, 56182575, 56994075, 35696286, 22278997, 22278998, 22278999, 264092, 264093, 264094, 264095, 264259, 29331822, 29331824, 56182181, 66714117, 29331825, 29331826, 60432289, 29331827, 29331828, 35696052, 33656970, 264105, 264508, 264905, 264906, 264907, 264908, 29331830, 66712502, 52644045, 56182435, 265007, 265009, 264910, 60170831, 264592, 60431735, 60433356, 33657402, 264757, 60433438, 55812038, 264758, 21906754, 52646317, 33109954, 52644296, 87168474, 265011, 87168559, 264601, 265017, 265018, 264604, 265019, 264448, 264369, 264288, 264766, 52644229, 21906768, 21908767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 60170615, 52644150, 33657023, 65274620, 33657109, 27486261, 27486264, 33657349, 35695763, 264628, 263972, 18108374, 55810764, 35696423, 55811576, 65274791, 35695855, 60431850, 264636, 52644332, 56182323, 60170394, 83373044, 18108385, 18108387, 18108388, 56526486, 87168518, 60432113, 22279000, 22279002, 264482, 264554, 264486
89	80077389 (177, 178)	Novel Protein sim. GBank gjl1710383[sp P46352 RIPX_BACSU - PROBABLE INTEGRASE/RECOMBINASE RIPX		264600
90	82115999 (179, 180)	Novel Protein sim. GBank gjl2498891[sp P76403 YEGQ_ECOLI - PUTATIVE PROTEASE IN BAER-OGRK INTERGENIC REGION	UNCLASSIFIED	264760
91	78906950 (181, 182)	Novel Protein sim. GBank gjl3367754[emb CAA20079] - (AL031155) hypothetical protein SC3A7.16c [Streptomyces coelicolor]	protease	265006
92	79554871 (183, 184)	Novel Protein sim. GBank gjl2895095 (AF011337) - putative E1-E2 ATPase [Mus musculus]	UNCLASSIFIED	264691
93	80496778 (185, 186)	Novel Protein sim. GBank gjl171919[sp P46920 OPUA_BACSU - GLYCINE BETAINES TRANSPORT ATP-BINDING PROTEIN OPUAA	ATPase_associated	264907, 264908, 264910, 265009, 264605, 264769
94	79846649 (187, 188)	Novel Protein sim. GBank gjl171919[sp P46920 OPUA_BACSU - GLYCINE BETAINES TRANSPORT ATP-BINDING PROTEIN OPUAA	transport	264906
95	11080238 (189, 190)			264594

96	94322125 (191, 182)	Novel Protein sim. GBank gij4589560[dbj BAA16802.1] - (AB023175) KIAA0958 protein [Homo sapiens]		UNCLASSIFIED	22278995, 22278999, 264259, 29331822, 29331826, 35696052, 29146499, 264509, 264906, 264907, 264908, 264909, 265007, 265008, 264910, 265009, 264593, 265010, 265017, 264604, 265019, 18108351, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264692, 33657182, 35695763, 264628, 264629, 18108379, 264631, 264636, 18108381, 264559, 18108382, 83373044, 22278902, 264508
97	79605200 (193, 194)	Novel Protein sim. GBank gij4583559[emb CAB40388.1] - (AJ005255) OxyR [Erwinia chrysanthemi]		UNCLASSIFIED	
98	79427000 (195, 196)	Novel Protein sim. GBank gij1001693[dbj BAA10430] - (D84002) hypothetical protein [Synechocystis sp.]		UNCLASSIFIED	264909
99	20466524 (197, 198)	Novel Protein sim. GBank gij1169479[sp P43925 EFG_HAEIN - ELONGATION FACTOR G (EF-G)]		UNCLASSIFIED	264605
100	79840113 (199, 200)	Novel Protein sim. GBank gij480897[pir S37485 - gene msg1 protein - mouse]		UNCLASSIFIED	264693
101	80203298 (201, 202)	Novel Protein sim. GBank gij2894166[emb CAA11773.1] - (AJ223998) PCZA361.18 [Amycolatopsis orientalis]		UNCLASSIFIED	265020, 264102, 263972
102	20467259 (203, 204)	Novel Protein sim. GBank gij1731040[sp P54509 YQHJL - BACSU - HYPOTHETICAL HELICASE IN SINI-GCVT INTERGENIC REGION]		synthase	264605
103	20466368 (205, 206)	Novel Protein sim. GBank gij1731040[sp P54509 YQHJL - BACSU - HYPOTHETICAL HELICASE IN SINI-GCVT INTERGENIC REGION]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	264605
104	80247572 (207, 208)	Novel Protein sim. GBank gij1854065[emb CAA58337] - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264591, 264595, 264602
105	79605206 (209, 210)	Novel Protein sim. GBank gij1685117 (U70770) - furrowed [Drosophila melanogaster]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complement	264508
106	28382058 (211, 212)	Novel Protein sim. GBank gij1705505[sp P54729 BS4 - MOUSE - BS4 PROTEIN]	Contains protein domain (PF00627) - UBA domain	UNCLASSIFIED	264511, 265009
107	80057791 (213, 214)	Novel Protein sim. GBank gij4887229[gb AAD32244.1 AF150755] - (AF150755) microtubule-actin crosslinking factor [Mus musculus]		ATPase-associated	29331824, 264591, 21906754, 265019
108	80237836 (215, 216)	Novel Protein sim. GBank gij263577[emb CAB15264] - (Z99120) similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]	Contains protein domain (PF00005) - ABC transporter	transport	18108374, 35695917, 22278996, 264113, 264600, 264602, 264603, 265017, 264910, 264906, 264636, 264766
109	95194148 (217, 218)	Novel Protein sim. GBank gij233079[emb CAB11265] - (Z98601) carboxypeptidase s precursor [Schizosaccharomyces pombe]		UNCLASSIFIED	264758, 264603, 264630, 264636, 264637
110	79582823 (219, 220)				264687
111	39565458 (221, 222)				264564
112	7956038 (223, 224)			UNCLASSIFIED	264908
113	17959439 (225, 226)			UNCLASSIFIED	265007
114	80502101 (227, 228)			UNCLASSIFIED	264769

115	80251003 (229, 230)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 73. contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	52645156, 52645080, 33656970, 264592, 21906754, 27486264, 18108379, 35696423, 264635, 52644332, 18108382
116	81298689 (231, 232)				264905, 264906, 264907, 264908, 264909, 264910, 264758, 265010, 264763, 264682, 264764, 264766, 264685, 264686, 264768, 264769, 33657023, 264693, 33657109, 264628, 18108374, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 56526486, 264565, 264566
117	79636695 (233, 234)		Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	264639, 264693
118	80222170 (235, 236)				263974
119	91013071 (237, 238)	Novel Protein sim. GBank gij732526 (U22327) - alpha2(IV) collagen [Caenorhabditis elegans]		UNCLASSIFIED	22278996, 29331824, 60432289, 265007, 60433438, 264603, 264605, 18108351, 264769, 264689, 265020, 264534, 27486261, 264558, 83373044, 18108385, 264564
120	8756491 (239, 240)	Novel Protein sim. GBank gij2131219 [pir]S50157 - cyclin-dependent kinase chain SRB10 - yeast (Saccharomyces cerevisiae)		kinase	264603
121	80026153 (241, 242)				264595
122	20457620 (243, 244)	Novel Protein sim. GBank gij2052147 [emb]CAB081371 - (Z94752) ksgA [Mycobacterium tuberculosis]	Contains protein domain (PF00398) - Ribosomal RNA adenine dimethylases	transferase	264605
123	8758278 (245, 246)				264604
124	79104017 (247, 248)	Novel Protein sim. GBank gij2833385 [sp]Q43134 [UGST_SORBI - GRANULE-BOUND GLYCOPROTEIN (STARCH) SYNTHASE PRECURSOR		synthase	18108394, 18108397, 265006, 265007, 265008, 265010, 265011, 18108355, 18108379, 18108380, 18108384
125	87797986 (249, 250)	Novel Protein sim. GBank gij475542 (U08255) - glutamate receptor delta-1 subunit [Rattus norvegicus]	Contains protein domain (PF00060) - Ligand-gated ion channel	misc_channel	264508, 264906, 265009, 264596, 22279002
126	56701283 (251, 252)	Novel Protein sim. GBank gij5102785 [emb]CAB45200.1 - (AL079308) putative transcriptional regulator [Streptomyces coelicolor]			264511
127	20467267 (253, 254)				264605
128	80248473 (255, 256)	Novel Protein sim. GBank gij130120 [sp]P23620 [PHOB_PSEAE - PHOSPHATE REGULON TRANSCRIPTIONAL REGULATORY PROTEIN PHOB	Contains protein domain (PF00072) - Response regulator receiver domain	phosphatase	264907, 264909, 264910, 264600, 264601, 264603, 264605, 18108351, 264693, 264557
129	85290543 (257, 258)	Novel Protein sim. GBank gij2508493 [sp]P38036 [YGCB_ECOLI - HYPOTHETICAL 100.5 KD PROTEIN IN IAP-CYSH INTERGENIC REGION (X83413) U88 [Human herpesvirus 6]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	UNCLASSIFIED	35686423, 35695855, 264600, 264602, 264603, 264604, 264605, 264508, 264906, 264564, 264628, 264682, 264565, 264683
130	80085583 (259, 260)	Novel Protein sim. GBank gij854055 [emb]CAA583371 - (X83413) U88 [Human herpesvirus 6]			264634
131	94995022 (261, 262)	Novel Protein sim. GBank gij1076038 [pir]S54860 - ABC transporter PstC-2 chain - Mycobacterium tuberculosis	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	18108376, 264769, 29331826, 264689, 22278996, 265021, 264600, 264511, 264601, 264602, 264605, 264905, 264638

132	10687692 (263, 264)	Novel Protein sim. GBank gj1877340[emb]CAB07058] - (Z92771) accA3 [Mycobacterium tuberculosis]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSase)	carboxylase	264636 264905, 264689
133	04630883 (265, 266)	Novel Protein sim. GBank gj4585838[emb]CAB40932.1] - (AL049630) putative NADH dehydrogenase [Streptomyces coelicolor]		dehydrogenase	264905, 264605, 265021
134	79834660 (267, 268)	Novel Protein sim. GBank gj1460074[emb]CAB01049] - (Z77250) hypothetical protein Rv2566 [Mycobacterium tuberculosis]			264634
135	19885057 (269, 270)	Novel Protein sim. GBank gj2125896[emb]CAA73511] - (Y13070) isopolyglutamate synthase [Streptomyces coelicolor]		synthase	264508
136	79846083 (271, 272)				264683, 264685, 264686, 264691, 264692, 264693
137	78619770 (273, 274)	Novel Protein sim. GBank gj15420387[emb]CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	18108374, 18108385, 33657109, 33657182, 265010, 22278998, 265006, 265007, 265008, 265009, 264693
138	78635971 (275, 276)	Novel Protein sim. GBank gj15689912[emb]CAB52075.1] - (AL109732) putative mutase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01817) - Chorismate mutase	dehydrogenase	22278996, 265007, 264910, 60433356, 265010, 264602, 264605, 264766, 264688, 264769, 264693, 32833986, 18108374, 18108387
139	86688076 (277, 278)			UNCLASSIFIED	264908
140	79825759 (279, 280)	Novel Protein sim. GBank gj3581916[emb]CAA20855] - (AL031545) mutS family DNA mismatch repair protein [Schizosaccharomyces pombe]		nuclease	264600 264602, 265017
141	20700094 (281, 282)				
142	80028104 (283, 284)	Novel Protein sim. GBank gj3334127[sp]P97303[BAC2, MOUSE - TRANSCRIPTION REGULATOR PROTEIN BACH2 (BTB AND CNC HOMOLOG 2)]		UNCLASSIFIED	264600
143	11072274 (285, 286)				263978, 264600, 264910, 264632, 264508, 264553, 264564, 264591, 264556, 264508, 264629, 264639
144	95009102 (287, 288)	Novel Protein sim. GBank gj140807[sp]P24536[Y121, BURCE - INSERTION EVIDENCE=genefinder:084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]		UNCLASSIFIED	22278996, 264602
145	80027058 (289, 290)	Novel Protein sim. GBank gj140807[sp]P24536[Y121, BURCE - INSERTION EVIDENCE=genefinder:084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]	Contains protein domain (PF01675) - Transposase		264687
146	13085662 (291, 292)	Novel Protein sim. GBank gj140807[sp]P24536[Y121, BURCE - INSERTION EVIDENCE=genefinder:084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]			264905, 264906, 264909, 264510, 265009, 60433356, 264600, 264601, 264604, 264605, 264687, 264769, 18108385, 65274791, 18108387
147	94320366 (293, 294)	Novel Protein sim. GBank gj2827608[emb]CAA16663] - (AL021646) uvrD2 [Mycobacterium tuberculosis]		helicase	
148	80248804 (295, 296)	Novel Protein sim. GBank gj2916947[emb]CAA17585] - (AL021999) hypothetical protein Rv0986 [Mycobacterium tuberculosis]		transport	265009, 265010, 264600, 264602, 264603, 264604, 264605, 264693, 33657109, 264636

149	80249373 (297, 298)	Novel Protein sim. GBank gij1723073[sp Q11040 Y081_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.01 (A)224340] maltose phosphorylase [Lactobacillus sanfrancisco]	Contains protein domain (PF00005) - ABC transporter	265010, 264600, 264601, 264603, 264604, 27486265, 264636
150	20294748 (299, 300)	Novel Protein sim. GBank gij3724125[emb CAA11905] - (A)224340] maltose phosphorylase [Lactobacillus sanfrancisco]		264600
151	20726398 (301, 302)	Novel Protein sim. GBank gij729312[sp P07651 DEOB_ECOLI - PHOSPHOPENTOMUTASE (PHOSPHODEOXYRIBOMUTASE)	Contains protein domain (PF01676) - Metalloenzyme superfamily	264602
152	95002877 (303, 304)	Novel Protein sim. GBank gij2497952[sp P55667 Y4TM_RHISN - HYPOTHETICAL HYDROLASE/PEPTIDASE Y4TM	peptidase	264602
153	80256665 (305, 306)	Novel Protein sim. GBank gij3123021[sp Q90508 VIT1_FUNHE - VITELLOGENIN I PRECURSOR (VTG I) (CONTAINS: LIPOVITELLIN 1 (LV1); PHOSVITIN (PV); LIPOVITELLIN 2 (LV2))	UNCLASSIFIED	264593
154	82305966 (307, 308)	Novel Protein sim. GBank gij418697[pir JUN0443 - transcription initiation factor sigma homolog hrdB - Streptomyces aureofaciens	Contains protein domain (PF00140) - Sigma-70 factor	264910, 264762, 264691, 264634
155	20429859 (309, 310)	Novel Protein sim. GBank gij628710[pir JSA1739 - hypotheical protein - Escherichia coli	UNCLASSIFIED	264565
156	39564742 (311, 312)	Novel Protein sim. GBank gij3695013 (AF052586) - CtrA [Pseudomonas aeruginosa]	Contains protein domain (PF00142) - 4Fe-4S iron sulfur cluster binding proteins. NifH/HnfX family	264691
157	10358887 (313, 314)	Novel Protein sim. GBank gij1073072[pir JCS5543 - cmaU protein - Pseudomonas syringae pv. syringae	UNCLASSIFIED	264905
158	79761838 (315, 316)	Novel Protein sim. GBank gij1073072[pir JCS5543 - cmaU protein - Pseudomonas syringae pv. syringae	UNCLASSIFIED	265008
159	78890378 (317, 318)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00400) - WD domain, G-beta repeat	264605
160	11075119 (319, 320)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00327) - Ribosomal protein L30p/L7e	22278996, 264600, 264603, 35695917, 32833986, 35696423, 264636
161	80055007 (321, 322)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	264112, 264532, 22279002
162	80016371 (323, 324)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	264639
163	11692306 (325, 326)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	264905, 284907, 264600
164	80077902 (327, 328)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	264691
165	10856067 (329, 330)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	264605, 264486
166	88095003 (331, 332)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	265010
167	16395460 (333, 334)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	264600
168	80078362 (335, 336)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	UNCLASSIFIED	264556, 264557, 264558, 264559
169	80239581 (337, 338)	Novel Protein sim. GBank gij1173023[sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30		

170	79512364 (339, 340)	Novel Protein sim. GBank gij140888[sp]P27847[YIGK_ECOLI - HYPOTHETICAL 15.4 KD PROTEIN IN RECQ-PLDB INTERGENIC REGION (F138)]	Contains protein domain (PF01810) - LysE type translocator		264908
171	95293073 (341, 342)				264595, 264604
172	37787007 (343, 344)	Novel Protein sim. GBank gij4210905[gb]AAD12048.1] - (AF045609) AgIG [Sinorhizobium meliloti]	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	264769
173	57529660 (345, 346)	Novel Protein sim. GBank gij132854[sp]P02387[RL2_ECOLI] - 50S RIBOSOMAL PROTEIN L2	Contains protein domain (PF00181) - Ribosomal Proteins L2	ribosomalprot	264769
174	95293078 (347, 348)	Novel Protein sim. GBank gij1881350[gb]BAA19371] - (AB001488) PROBABLE TRANSPORT PROTEIN, SIMILAR TO ANTIBIOTIC TRANSPORT-ASSOCIATED PROTEIN ACTII IN STREPTOMYCES COELICOLOR. [Bacillus subtilis]		transport	264510, 264593, 264602, 264603, 264605, 264762, 264693
175	79758270 (349, 350)	Novel Protein sim. GBank gij207272[emb]CAB08326] - (Z95121) manA [Mycobacterium tuberculosis]		isomerase	264565
176	80066896 (351, 352)	Novel Protein sim. GBank gij1055198 (J40187) - similar to PIR:A41724 chicken LD (limb deformity) gene product and to formin; also P-rich region similar to collagen [Caenorhabditis elegans]		UNCLASSIFIED	264907, 264910, 264681, 264558
177	86684852 (353, 354)	Novel Protein sim. GBank gij232673[emb]CAB10952] - (Z98268) hypothetical protein Rv1695 [Mycobacterium tuberculosis]	Contains protein domain (PF01513) - Domain of unknown function	UNCLASSIFIED	264768, 60424179, 264687, 264688, 264769, 29331826, 60432289, 18108376, 264689, 18108387, 32833986, 22278996, 265020, 264600, 264601, 264602, 264603, 264604, 264605, 264635, 264762, 264636, 264908, 264564, 264637, 264638, 264486, 60433356, 264766
178	78559526 (355, 356)	Novel Protein sim. GBank gij1906596 (J81788) - kinesin-73 [Drosophila melanogaster]		struct	264693, 33657109, 264635
179	20263112 (357, 358)			UNCLASSIFIED	264563
180	80488958 (359, 360)	Novel Protein sim. GBank gij1169367[sp]P45256[DNAB_HAEIN - REPLICATIVE DNA HELICASE		helicase	264769
181	78585369 (361, 362)	Novel Protein sim. GBank gij3170615 (AF059485) - DOG4 [Mus musculus]		UNCLASSIFIED	21906767, 264635, 264639, 18108384
182	80577899 (363, 364)				
183	11614017 (365, 366)	Novel Protein sim. GBank gij1076627[pir]S54172 - inorganic pyrophosphatase (EC 3.6.1.1) - common tobacco		UNCLASSIFIED	264259, 35696052, 56182435, 264511, 265018, 33657109, 264555, 264568
184	10174167 (367, 368)	Novel Protein sim. GBank gij4371280[gb]AAD18138] - (AC008260) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264690

185	21600822 (369, 370)	Novel Protein sim. GBank gll3006178[emb]CAA18398.1] - (AL022304) putative mma transport regulator [Schizosaccharomyces pombe]		UNCLASSIFIED	264604	
186	60070329 (371, 372)	Novel Protein sim. GBank gll282802[sp]P94408[YCLF_BACSU - HYPOTHETICAL 53.3 KD PROTEIN IN SFP-GERKA INTERGENIC REGION		transport	264595	
187	80186611 (373, 374)	Novel Protein sim. GBank gll3150260[emb]CAA19179] - (AL023634) cydlin [Schizosaccharomyces pombe]		UNCLASSIFIED	264369	
188	20464942 (375, 376)	Novel Protein sim. GBank gll2145853[pi]S72938 - hix protein - Mycobacterium leprae		kinase	264605	
189	82338215 (377, 378)	Novel Protein sim. GBank gll1881244[db]BAA19271] - (AB001488) SIMILAR TO PYRUVATE OXIDASE AND ACETOLACTATE SYNTHASE. [Bacillus subtilis]		UNCLASSIFIED	35696052, 264602, 264605, 264762, 264689, 35695917, 18108370, 18108372, 264638,	
190	80066821 (379, 380)	Novel Protein sim. GBank gll120226[sp]P28725[FKBP_STRCH - FK506-BINDING PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PIPIASE) (ROTAMASE)	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	synthase	264565	
191	86095012 (381, 382)	Novel Protein sim. GBank gll1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)	Contains protein domain (PF00254) - isomerase	isomerase	264508, 264604, 264605, 264769, 264555	
192	16333379 (383, 384)	Novel Protein sim. GBank gll4980892[gb]AAD35474.1[AE00171] - (AE001718) ABC transporter, ATP-binding protein [Thermotoga maritima]			264567	
193	78810127 (385, 386)	Novel Protein sim. GBank gll1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)			264908, 264693	
194	20464949 (387, 388)	Novel Protein sim. GBank gll1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)			264605	
195	13518389 (389, 390)	Novel Protein sim. GBank gll1705461[sp]P53656[BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7.8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)		transport	264636	
196	95005569 (391, 392)	Novel Protein sim. GBank gll312305[sp]Q27778[K6PF_SCHMA - 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE) (PHOSPHOHEXOKINASE)	Contains protein domain (PF00202) - gaba Aminotransferases class-III pyridoxal phosphate		264600, 264669, 264638	
197	80248665 (393, 394)	Novel Protein sim. GBank gll312305[sp]Q27778[K6PF_SCHMA - 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE) (PHOSPHOHEXOKINASE)	Contains protein domain (PF00365) - kinase	kinase	264602, 264682, 264692, 18108374	
198	78163635 (395, 396)	Novel Protein sim. GBank gll1781203[emb]CA806110] - (Z83859) gnd [Mycobacterium tuberculosis]			264636	
199	78890715 (397, 398)	Novel Protein sim. GBank gll1781203[emb]CA806110] - (Z83859) gnd [Mycobacterium tuberculosis]	Contains protein domain (PF00393) - 6-phosphogluconate dehydrogenases		265008	
200	78413849 (399, 400)	Novel Protein sim. GBank gll2842222 [AF030885] - telomere-associated recQ-like helicase [Usiliago maydis]		UNCLASSIFIED	264595, 264596	
201	86945924 (401, 402)	Novel Protein sim. GBank gll2894379[emb]CAA74911.1] - (Y14573) ring finger protein [Hordeum vulgare]		UNCLASSIFIED	29331826, 265007, 264512, 33657402, 264596, 265017, 18108351, 264682, 264683, 264767, 264629, 55810764, 264634, 264635, 58182323, 60432113, 22278000	



202	79568046 (403, 404)	Novel Protein sim. GBank gi 231772 sp P30598 CHS1_1 - CHITIN SYNTHASE 1 (CHITIN-UDP ACETYL-GLUCOSAMINYL TRANSFERASE 1)	Contains protein domain (PF01644) - Chitin synthase	synthase	264600	
203	79843927 (405, 406)	Novel Protein sim. GBank gi 1504042 dbj BAA132201 - (D86984) similar to yeast adenylate cyclase (S56776) [Homo sapiens]			22278995, 29331822, 29331825, 28331827, 264906, 21906754, 264683, 21906766, 21906769, 35696423, 264556	
204	79855188 (407, 408)			UNCLASSIFIED	264909	
205	10080583 (409, 410)	Novel Protein sim. GBank gi 2633808 emb CAB133101 - (Z99111) similar to hypothetical proteins [Bacillus subtilis]		transport	264909	
206	8758473 (411, 412)			UNCLASSIFIED	264604	
207	20754522 (413, 414)	Novel Protein sim. GBank gi 2134381 pir S60678 - polybromo 1 protein - chicken		UNCLASSIFIED	264556	
208	20289261 (415, 416)				264605	
209	80071069 (417, 418)	Novel Protein sim. GBank gi 2501040 sp O5814 SYP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE-TRNA LIGASE) (PRORS)			264605, 264689	
210	50168800 (419, 420)				264905, 264907, 264909, 264766, 264687, 264691, 264629, 18108374, 264638	
211	80034539 (421, 422)				263978	
212	82442474 (423, 424)	Novel Protein sim. GBank gi 5031809 ref NP_005536.1 pISLR - immunoglobulin superfamily containing leucine-rich repeat		UNCLASSIFIED	264508, 264905, 264906, 264907, 264908, 264600, 264762, 264534, 264632, 264634, 264635, 264639, 264486	
213	80249562 (425, 426)	Novel Protein sim. GBank gi 3123258 sp O331231 LEU2_MYCLE - 3- ISOPROPYLMALATE DEHYDRATASE LARGE SUBUNIT (ISOPROPYLMALATE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMI)	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	isomerase	22278996, 264508, 264600, 264602, 264603, 264605, 33657023, 264565, 264486	
214	80079381 (427, 428)	Novel Protein sim. GBank gi 116236 sp P19421 CH60_COXBU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN B)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	264600, 264693	
215	14973283 (429, 430)					
216	80177716 (431, 432)	Novel Protein sim. GBank gi 3417297 (AC002310) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264629	
217	79603634 (433, 434)	Novel Protein sim. GBank gi 2506924 sp P49754 VP41_HUMAN - VACUOLAR ASSEMBLY PROTEIN VPS41 HOMOLOG (S53)	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_rna_bind	264448	
218	80258475 (435, 436)	Novel Protein sim. GBank gi 1173288 sp P38106 RSEA_ECOLI - SIGMA-E FACTOR NEGATIVE REGULATORY PROTEIN			264508	
219	20438787 (437, 438)	Novel Protein sim. GBank gi 1781097 emb CAB062311 - (Z83864) gltB [Mycobacterium tuberculosis]		mapolymerase	264594	
220	13489572 (439, 440)	Novel Protein sim. GBank gi 2984703 (AF052427) - unknown [Trypanosoma cruzi]		synthase	264604	
221	11287488 (441, 442)	Novel Protein sim. GBank gi 4587313 dbj BAA176709.1 - (AB025248) alpha-1,2-mannosidase [Bacillus sp. M-90]		nucleaseinhib	264689	
				UNCLASSIFIED	264555	

222	79862802 (443, 444)	Novel Protein sim. GBank gij1877268[emb]CAB07049] - (Z92770) hypothetical protein Rv0143c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 264769, 35696423
223	83053869 (445, 446)			UNCLASSIFIED	264906, 264907, 264603
224	79557920 (447, 448)			UNCLASSIFIED	264684, 264693
225	79559541 (449, 450)	Novel Protein sim. GBank gij2274851[dbj]BAA21515] - (D84159) 3-7 gene product [Homo sapiens]		UNCLASSIFIED	264692
226	79172397 (451, 452)	Novel Protein sim. GBank gij868245 (U29488) - C56C10.7 gene product [Caenorhabditis elegans]		UNCLASSIFIED	22278998, 264112, 33657023, 263981
227	81777196 (453, 454)			UNCLASSIFIED	35695917, 264636, 264907
228	79872285 (455, 456)				264768, 264907, 264908, 264692, 264593, 264639
229	79838268 (457, 458)				264908, 264910
230	11013209 (459, 460)			UNCLASSIFIED	264631
231	20622207 (461, 462)	Novel Protein sim. GBank gij1835114[emb]CAA71733] - (Y10744) homoserine O-acetyltransferase [Leptospira meyeri]			264906, 264600, 264603, 264692
232	80055035 (463, 464)			UNCLASSIFIED	264600, 264603, 264605, 264687, 264769
233	80063054 (465, 466)	Novel Protein sim. GBank gij2842340 (AF032970) - imidazole propionate hydrolase [Pseudomonas putida]	Contains protein domain (PF00449) - Urease		264604
234	7523998 (467, 468)	Novel Protein sim. GBank gij3510505 (AF030881) - polypeptide [Fugu rubripes]		UNCLASSIFIED	264369
235	80203671 (469, 470)	Novel Protein sim. GBank gij2104609[emb]CAB08805] - (Z95398) PckA [Mycobacterium leprae]		UNCLASSIFIED	264106
236	78940001 (471, 472)			carboxylase	264905
237	11755273 (473, 474)				264681
238	79461401 (475, 476)				264639
239	82435190 (477, 478)	Novel Protein sim. GBank gij2495617[sp]Q57252[YDJ]_HAEIN - HYPOTHETICAL PROTEIN HI1163	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.	UNCLASSIFIED	264906, 265010, 264603, 264762, 264682, 264638, 264638, 264486
240	21635575 (479, 480)	Novel Protein sim. GBank gij183458[sp]P75796[YLIA]_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLIA	transport		264259, 264769
241	80377307 (481, 482)	Novel Protein sim. GBank gij3875920[emb]CAB04111] - (Z81503) predicted using GeneFinder; similar to collagen; cDNA EST EMBL:D65450 comes from this gene; cDNA EST EMBL:D68888 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	264908, 264909, 264764, 264639
242	82148454 (483, 484)			UNCLASSIFIED	264489, 264907, 264908, 264511, 264760, 264764, 264892, 264635, 264637
243	79633207 (485, 486)			UNCLASSIFIED	264906
244	80246682 (487, 488)	Novel Protein sim. GBank gij2624302[emb]CAA15575] - (AL008987) aid [Mycobacterium tuberculosis]	dehydrogenase		264600, 264602, 264605, 264769, 264689
245	79863543 (489, 490)	Novel Protein sim. GBank gij2920625 (AF044499) - vgrE protein [Escherichia coli]		UNCLASSIFIED	264907, 264758
246	79162929 (491, 492)	Novel Protein sim. GBank gij5420387[emb]CAB46679.1] - (A1243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF01106) - NIFI-like domain		264637, 18108381, 18108387, 264565

247	79873185 (493, 494)	Novel Protein sim. GBank gij1839006[emb]CAB06648] - (ZB5982) argB [Mycobacterium tuberculosis]		kinase	264909, 264691, 35696423, 18108387
248	80488983 (495, 496)	Novel Protein sim. GBank gij1168574[sp]P42464[ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	35996286, 264907, 264511, 264602, 264768, 264688, 265021, 35695855, 18108385
249	79764845 (497, 498)			UNCLASSIFIED	264907, 264910, 265011, 264762, 264636
250	79619980 (499, 500)				21908768, 264692
251	84359489 (501, 502)			UNCLASSIFIED	52645156, 29331822, 29331824, 52644045, 265018, 21906765, 21906768, 265020, 27486261, 27486265, 35695763, 18108376, 264556, 264559, 264565
252	79737758 (503, 504)	Novel Protein sim. GBank gij3327166[db]BAA31651] - (AB014576) KIAA0676 protein [Homo sapiens]			264685, 264687, 264632
253	20443124 (505, 506)	Novel Protein sim. GBank gij3036880[emb]CAA18513] - (AL022374) putative ATP-dependent DNA helicase [Sireplomyces coelicolor]		helicase	264604
254	80027421 (507, 508)	Novel Protein sim. GBank gij3915488[sp]O34981[YJMB_BACSU - HYPOTHETICAL SYMPORTER IN COTT-RAPA INTERGENIC REGION (D17312) diarrheal toxin [Bacillus cereus]		UNCLASSIFIED	264508, 264806, 264602, 264687, 265021, 264486
255	11388315 (509, 510)	Novel Protein sim. GBank gij1665720[db]BAA04134] - (D17312) diarrheal toxin [Bacillus cereus]		UNCLASSIFIED	264593
256	80028158 (511, 512)	Novel Protein sim. GBank gij465787[sp]P34422[YL31_CAEEL - HYPOTHETICAL 86.0 KD PROTEIN F44B9.1 IN CHROMOSOME III	Contains protein domain (PF00326) - Prolyl oligopeptidase family	peptidase	264602, 264692
257	20289282 (513, 514)	Novel Protein sim. GBank gij1172039[sp]P42315[SCOA_BACSU - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE SUBUNIT A (SUCCINYL COA:3-OXOACID COA-TRANSFERASE) (OXCT A)	Contains protein domain (PF01144) - Coenzyme A transferase	transferase	264605
258	20459464 (515, 516)	Novel Protein sim. GBank gij3127836[emb]CAA18902] - (AL023496) hypothetical protein [Sireplomyces coelicolor]		UNCLASSIFIED	264604
259	79910152 (517, 518)			collagen	264681, 264686, 264692
260	20376437 (519, 520)			UNCLASSIFIED	264692, 264556
261	20285683 (521, 522)	Novel Protein sim. GBank gij123761[sp]P24221[HUTH_STRGR - HISTIDINE AMMONIA-LYASE (HISTIDASE)	Contains protein domain (PF00221) - Phenylalanine and histidine ammonia lyases	UNCLASSIFIED	264600
262	80189317 (523, 524)			UNCLASSIFIED	265017, 264369
263	88095045 (525, 526)	Novel Protein sim. GBank gij3924708[emb]CAA84646] - (Z35597) Weak similarity with sea squirt nidogen precursor protein (blastp score 71); cDNA EST EMBL: T02069 comes from this gene; cDNA EST EMBL: D76135 comes from this gene; cDNA EST EMBL: D73147 comes from this gene; cDNA EST EMB...		UNCLASSIFIED	264488, 264905, 264906, 264907, 264908, 264909, 264512, 264910, 264758, 264596, 264604, 265019, 264605, 264760, 18108351, 264763, 264764, 264288, 264766, 264768, 264769, 264691, 264692, 264693, 264628, 264634, 264635, 264555, 264636, 264638, 264639
264	87370826 (527, 528)	Novel Protein sim. GBank gij3043734[db]BAA25531] - (AB011177) KIAA0605 protein [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	protease	264259, 264908, 21906754, 265018, 265019, 265020

265	95355646 (528, 530)	Novel Protein sim. GBank gij4589624[dj]BAA76834.1] - (AB023207) KIAA0980 protein [Homo sapiens]		kinase	264488, 35696286, 29331824, 56182181, 35696052, 264508, 264905, 264906, 264907, 66712502, 264908, 264909, 264511, 264512, 264910, 264592, 264595, 264758, 264596, 55811386, 264600, 265017, 264603, 264604, 264605, 264760, 18108351, 264762, 264681, 264764, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264691, 33657023, 33657109, 33657182, 264628, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264555, 264638, 83373044, 56526486, 87168518, 264564, 264566, 264486, 264600
266	79588075 (531, 532)			UNCLASSIFIED	264828
267	11362222 (533, 534)			UNCLASSIFIED	264687, 264769, 264689
268	79509566 (535, 536)			UNCLASSIFIED	264602
269	80025810 (537, 538)			UNCLASSIFIED	264693
270	84361144 (539, 540)	Novel Protein sim. GBank gij4507367[ref]NP_003182.1 pTARS - threonyl-tRNA synthetase		UNCLASSIFIED	264909, 264693
271	79552301 (541, 542)	Novel Protein sim. GBank gij4980738 AAD35331.1 AE00170 - (AE001707) glucose-1-phosphate adenylyltransferase [Thermotoga maritima]		UNCLASSIFIED	264908
272	9874778 (543, 544)	Novel Protein sim. GBank gij1188224 sp P44569 5NTD_HAEIN - PROBABLE 5'-NUCLEOTIDASE PRECURSOR		UNCLASSIFIED	264688
273	12840694 (545, 546)			UNCLASSIFIED	264564
274	39524246 (547, 548)	Novel Protein sim. GBank gij3253159 (AF005355) - Translation initiation factor eIF2C [Oryctolagus cuniculus]		UNCLASSIFIED	264907, 264908, 264909, 264766, 264768, 264691, 264632, 264638
275	82787041 (549, 550)	Novel Protein sim. GBank gij134920 sp P21997 SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)			265008, 60432229
276	86671073 (551, 552)	Novel Protein sim. GBank gij128021 sp P20964 OBG_BACSU - SPOB-ASSOCIATED GTP-BINDING PROTEIN		ribosomalprot	264600, 18108387
277	60078735 (553, 554)			UNCLASSIFIED	264689
278	12866947 (555, 556)	Novel Protein sim. GBank gij79839 pir S03812 - uvrB protein - Micrococcus luteus		nuclease	264508, 264604, 21906764, 264638, 264557, 264404
279	95292719 (557, 558)				264259
280	5603617 (559, 560)	Novel Protein sim. GBank gij312160 sp Q18964 YLN2_CAEEL - HYPOTHETICAL 46.2 KD TRP-ASP REPEATS CONTAINING PROTEIN D2013.2 IN CHROMOSOME II			18108392, 264634, 264555, 264556, 264557, 264558
281	80249599 (561, 562)			UNCLASSIFIED	265019
282	18598682 (563, 564)			UNCLASSIFIED	264555
283	20614211 (565, 566)				

284	91212160 (567, 568)	Novel Protein sim. GBank gij2428094 (U56632) - acetyl xylan esterase; AxeA [Thermotoga neapolitana]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	UNCLASSIFIED	35696052, 29331828, 264508, 264905, 264600, 264602, 264605, 264682, 264764, 56181562, 21906764, 18108376, 264636, 264559, 18108387
285	6757940 (569, 570)			UNCLASSIFIED	264603
286	80503235 (571, 572)	Novel Protein sim. GBank gij2072674[emb]CAB08305] - (Z95120) rhlE [Mycobacterium tuberculosis]	Contains protein domain (PF00270) - ATPase-associated DEAD/DEAH box helicase	UNCLASSIFIED	35696052, 264769, 264638
287	12745521 (573, 574)			UNCLASSIFIED	264689
288	20756502 (575, 576)	Novel Protein sim. GBank gij7655323[bs]157676 - (S74439) silk fibroin heavy chain (C-terminal) [Bombyx mori-silkworms, Peptide Partial, 633 aa] [Bombyx mori]	collagen	UNCLASSIFIED	264557
289	80043804 (577, 578)	Novel Protein sim. GBank gij1870009[emb]CAB06860] - (Z92539) hypothetical protein Rv1019 [Mycobacterium tuberculosis]	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	UNCLASSIFIED	264593, 264600
290	50430175 (579, 580)			UNCLASSIFIED	264768
291	20747431 (581, 582)	Novel Protein sim. GBank gij2506664[sp]P40120[YDCG, ECOLI - 59.4 PROTEIN IN TRG-RIML INTERGENIC REGION PRECURSOR		UNCLASSIFIED	264601
292	80052555 (583, 584)	Novel Protein sim. GBank gij625182 (L39015) - mitochondrial glutamyl-tRNA synthetase [Saccharomyces cerevisiae]		UNCLASSIFIED	264605
293	80062519 (585, 586)	Novel Protein sim. GBank gij1718069[sp]P53528[UVRD_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG	helicase	UNCLASSIFIED	264903, 264605, 264687, 264689, 264692
294	79830303 (587, 588)	Novel Protein sim. GBank gij117422[sp]P10040[CRB_DROME - CRUMBS PROTEIN PRECURSOR (95F)	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	35696052, 264906, 265011, 264628, 55811576
295	78444180 (589, 590)	Novel Protein sim. GBank gij1181619[db]BAA11565] - (D82364) a variant of TSC-22 [Gallus gallus]		UNCLASSIFIED	52644507, 29331822, 264592, 265020, 264639
296	79607076 (591, 592)	Novel Protein sim. GBank gij3649789[db]BAA33403] - (AB012226) SecA [Vibrio alginolyticus]	synthase	UNCLASSIFIED	264508
297	79631297 (593, 594)	Novel Protein sim. GBank gij5689967[emb]CAB52004.1] - (AL109663) putative membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264905, 264687, 264638
298	80418898 (595, 596)			UNCLASSIFIED	264905, 264691, 264639, 264766

299	95293298 (597, 598)	Novel Protein sim. GBank gij220637[dbjBAA014771] - (D10627) zinc finger protein [Mus musculus]	(Contains protein domain (PF00096) - Zinc finger, C2H2 type)		264488, 263994, 56994075, 22278997, 22278998, 22278999, 20281099, 29331824, 29331825, 29331826, 50432289, 29331827, 29331828, 264905, 264906, 264907, 264908, 52644045, 264909, 264511, 265008, 264910, 264595, 264596, 264758, 33657084, 87168559, 265018, 265019, 264764, 264288, 264766, 264687, 56181582, 264769, 21906765, 21906768, 21906769, 33657023, 264692, 33657109, 27486261, 18108370, 264628, 264629, 55811576, 35695855, 264631, 264634, 264635, 264638, 264639, 83373044, 18108387, 87168518, 22279000, 22279002, 264565, 264566, 264567
300	20711340 (599, 600)	Novel Protein sim. GBank gij145922 (M20981) - Iron dicitrate transport protein precursor [Escherichia coli]	UNCLASSIFIED	transport	264602
301	13511332 (601, 602)	Novel Protein sim. GBank gij1174661[spIP44594][TGT_HAEIN - QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA-GUANINE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME)]			264687
302	9875260 (603, 604)	Novel Protein sim. GBank gij1174661[spIP44594][TGT_HAEIN - QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA-GUANINE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME)]			264908
303	79574895 (605, 606)	Novel Protein sim. GBank gij67985[priHJUNVAV - helicase (EC 3.6.1.-) - Autographa californica nuclear polyhedrosis virus]		helicase	264689
304	20711344 (607, 608)	Novel Protein sim. GBank gij1174661[spIP44594][TGT_HAEIN - QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA-GUANINE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME)]			264602
305	80412520 (609, 610)	Novel Protein sim. GBank gij728867[spIP40602][APG_ARATH - ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR]			264763
306	8515878 (611, 612)	Novel Protein sim. GBank gij1657554[gbIAAB18082.1] - (U73857) hypothetical protein [Escherichia coli]	UNCLASSIFIED		263978
307	80222901 (613, 614)	Novel Protein sim. GBank gij1710612[spIQ10793][RNH2_MYCTU - PROBABLE RIBONUCLEASE HII (RNASE HII)]	UNCLASSIFIED		265010, 21906768, 265020, 18108374, 263977
308	80064305 (615, 616)	Novel Protein sim. GBank gij1710612[spIQ10793][RNH2_MYCTU - PROBABLE RIBONUCLEASE HII (RNASE HII)]	Contains protein domain (PF01351) - Ribonuclease HII	nuclease	264910, 264600, 264605, 264687, 264689, 264638, 18108387
309	80504138 (617, 618)	Novel Protein sim. GBank gij5420387[embI(CAB46678.1) - (AJ243459) proteophosphoglycan [Leishmania major]			264769
310	80053616 (619, 620)	Novel Protein sim. GBank gij11744522 (U34957) - phosphoribosylaminimidazole succinocarboxamide synthase [Mycobacterium tuberculosis]		synthase	264603
311	11090659 (621, 622)	Novel Protein sim. GBank gij11744522 (U34957) - phosphoribosylaminimidazole succinocarboxamide synthase [Mycobacterium tuberculosis]			264602
312	80054347 (623, 624)	Novel Protein sim. GBank gij11744522 (U34957) - phosphoribosylaminimidazole succinocarboxamide synthase [Mycobacterium tuberculosis]	UNCLASSIFIED		264566
313	80046168 (625, 626)	Novel Protein sim. GBank gij11744522 (U34957) - phosphoribosylaminimidazole succinocarboxamide synthase [Mycobacterium tuberculosis]			264603, 264567

314	87845112 (627, 628)	Novel Protein sim. GBank gij3661583 (AF092175) - ikaros [Danio rerio]	Contains protein domain (PF00320) - GATA zinc finger	dna_rna_bind	264259, 60432289, 29331828, 264905, 264906, 264908, 264909, 265008, 264910, 60432229, 33657402, 60433438, 33109954, 265011, 265017, 264603, 265018, 264288, 264766, 264692, 35695763, 264628, 264829, 264639, 60170394, 22279002, 264568
315	82356091 (629, 630)	Novel Protein sim. GBank gij1652620jbjBAA17540j - (D90907) pyridine nucleotide transhydrogenase beta subunit [Synecocystis sp.]			264508, 264600, 264762, 264687, 264768, 52644229, 264769, 264689, 264635, 264636, 264638, 264486
316	78911071 (631, 632)	Novel Protein sim. GBank gij118244spjP24176DAPE_ECOLI - SUCCINYL-DIAMINOPIMELATE DESUCCINYLAASE (SDAP)	UNCLASSIFIED	UNCLASSIFIED	264693
317	20466944 (633, 634)	Novel Protein sim. GBank gij4680228jbbAAD27583.1jAF11827 - (AF118274) DNB-5 [Homo sapiens]	Contains protein domain (PF00526) Dictyostellum (slime mold) repeats	transport	264908, 264909, 264910, 264593, 264594, 264760, 264288, 264768, 264769, 21906769, 264691, 264693, 264628, 65274791, 264635, 264636, 264638, 83373044, 22278002, 264566
318	94141836 (635, 636)	Novel Protein sim. GBank gij1149693jembjCAA60220j - (X86499) rbsC [Clostridium perfringens]	transport	transport	265018
319	17289360 (637, 638)	Novel Protein sim. GBank gij2811033spjO05314jGLGC_MYCTU - GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE (ADP-GLUCOSE SYNTHASE) (ADP-GLUCOSE PYROPHOSPHORYLAASE)	synthase	synthase	264687
320	13527675 (639, 640)	Novel Protein sim. GBank gij1680716 (U68234) - all-trans-retinoic acid 4-hydroxylase [Danio rerio]			
321	94134387 (641, 642)	Novel Protein sim. GBank gij1160355 (U33058) - UNC-89 [Caenorhabditis elegans]	cyto450	cyto450	264509, 264906, 264907, 264908, 265009, 264596, 264764, 264628, 264634, 264635, 264638, 264639, 83373044, 264567
322	68489053 (643, 644)	Novel Protein sim. GBank gij1160355 (U33058) - UNC-89 [Caenorhabditis elegans]	UNCLASSIFIED	UNCLASSIFIED	5581150, 264691, 60431528, 55810764
323	94853725 (645, 646)		UNCLASSIFIED	UNCLASSIFIED	264488, 265009, 264593, 264628, 264635
324	78174383 (647, 648)		UNCLASSIFIED	UNCLASSIFIED	264687
325	79862691 (649, 650)		UNCLASSIFIED	UNCLASSIFIED	264693
326	28774974 (651, 652)		UNCLASSIFIED	UNCLASSIFIED	264288, 18108385
327	79776267 (653, 654)	Novel Protein sim. GBank gij451544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]			264488, 264905, 264509, 264910
328	80253202 (655, 656)		UNCLASSIFIED	UNCLASSIFIED	264592
329	10173821 (657, 658)		UNCLASSIFIED	UNCLASSIFIED	264510
330	86597767 (659, 660)	Novel Protein sim. GBank gij4191358 (AF087825) - claudin-7 [Mus musculus]	UNCLASSIFIED	UNCLASSIFIED	264259, 264908
331	79754888 (661, 662)	Novel Protein sim. GBank gij80741jpirjS20912 - regulatory protein whB - Streptomyces coelicolor	transcriptfactor	transcriptfactor	264910, 264687, 264689, 264636, 264567
332	80071440 (663, 664)	Novel Protein sim. GBank gij114049spjP19480jAHPF_SALTY - ALKYL HYDROPEROXIDE REDUCTASE SUBUNIT F (ALKYL HYDROPEROXIDE REDUCTASE F52A PROTEIN)	reductase	reductase	35696423, 264636, 264638, 264565
333	13009555 (665, 666)				264687

334	80230771 (667, 668)	Novel Protein sim. GBank gjl322228 pir S32227 - glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - Corynebacterium glutamicum	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Va line dehydrogenase	dehydrogenase	264905, 264600, 264604, 264486
335	80057026 (668, 670)	Novel Protein sim. GBank gjl2193938 emb CAB09602  - (Z98800) glpQ2 [Mycobacterium tuberculosis]		esterase	264907, 264603, 264693, 18108374, 264636, 18108387
336	80414319 (671, 672)			UNCLASSIFIED	265009, 264766, 264686
337	11090829 (673, 674)				264602
338	95413134 (675, 676)	Novel Protein sim. GBank gjl5454074 ref NP_006303.1 pSMRT - silencing mediator for retinoid and thyroid hormone receptors	Contains protein domain (PF00249) - Myb-like DNA-binding domain	nucl_recpt	264569, 18108397, 22276998, 29331822, 20281099, 29331824, 58182181, 68714117, 29331825, 35696052, 29331828, 264508, 264509, 264905, 264906, 264807, 264908, 264909, 265006, 265008, 264910, 265009, 264758, 55812038, 65274444, 265011, 87168559, 265017, 265018, 265019, 264760, 55811150, 264681, 284762, 18108351, 264682, 264764, 264766, 264685, 264688, 264768, 52844229, 264693, 55811957, 35695917, 264692, 264693, 264628, 18108370, 18108374, 55811576, 35696423, 35695855, 264635, 264555, 264636, 264556, 264637, 264557, 18108380, 264638, 264558, 264639, 18108381, 83373044, 18108385, 87168518, 60432113
339	11398513 (677, 678)	Novel Protein sim. GBank gjl4001713 dbj BAA35087.1  - (AB015879) DnaK [Porphyromonas gingivalis]		eph	264593
340	80504149 (678, 680)	Novel Protein sim. GBank gjl2842699 sp Q92353 UBPC - PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C68.08 (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE) (DEUBIQUITINATING ENZYME)		ubiquitin	264905, 265019, 264768, 18108374
341	11075198 (681, 682)	Novel Protein sim. GBank gjl2688580 (AE001166) - conserved hypothetical protein [Borrelia burgdorferi]	Contains protein domain (PF00290) - Tryptophan synthase alpha chain	isomerase	264605
342	80054196 (683, 684)	Novel Protein sim. GBank gjl1684738 emb CAA7060.1  - (Y09452) Yed J hypothetical protein [Pseudomonas syringae]			264603, 264604
343	20466782 (685, 686)				264605
344	80428870 (687, 688)	Novel Protein sim. GBank gjl2117275 emb CAB09104  - (Z95618) hypothetical protein Rv0807 [Mycobacterium tuberculosis]		UNCLASSIFIED	264600, 264605, 264768, 18108370, 18108374, 35695855
345	80258853 (689, 690)	Novel Protein sim. GBank gjl3023317 sp Q48935 APHA_MYCRA - ACETYL POLYAMINE AMINOHYDROLASE		histone	264593
346	78831058 (691, 692)	Novel Protein sim. GBank gjl4239787 emb CAA75437  - (Y15166) NADP-glutamate dehydrogenase [Pseudomonas aeruginosa]	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Va line dehydrogenase	dehydrogenase	264905



347	79158195 (683, 684)	Novel Protein sim. GBank gij731675 sp P38795 YHN4_YEAST - HYPOTHETICAL 80.7 KD PROTEIN IN ERG7-NMD2 INTERGENIC REGION		UNCLASSIFIED	265006, 265008, 265010, 265018, 263967, 263981
348	80020208 (695, 696)	Novel Protein sim. GBank gij1073610 p S47672 - ugpB protein - Escherichia coli		transport	264602, 18108351, 18108387
349	17282112 (697, 698)	Novel Protein sim. GBank gij3261599 emb CA800917  - (Z77137) hypothetical protein Rv1277 [Mycobacterium tuberculosis]		nuclease	265007 265009, 264769, 264689, 18108370
350	80502370 (698, 700)	Novel Protein sim. GBank gij2959367 emb CAA17921  - (AL022117) hypothetical protein [Schizosaccharomyces pombe]		glycoprotein	264769, 264905, 264908
351	80501805 (701, 702)	Novel Protein sim. GBank gij4416302 gb AAD20307  - (AF105716) copia-type pol polyprotein [Zea mays]		protease	264595
352	11811585 (703, 704)	Novel Protein sim. GBank gij1174887 sp P42873 URE1_STAXY - UREASE ALPHA SUBUNIT (UREA AMIDOHYDROLASE)	Contains protein domain (PF00449) - Urease	UNCLASSIFIED	264604
353	80061653 (705, 706)	Novel Protein sim. GBank			
354	56626130 (707, 708)	Novel Protein sim. GBank		UNCLASSIFIED	264628
355	80046344 (709, 710)	Novel Protein sim. GBank	Contains protein domain (PF00072) - Response regulator receiver domain	UNCLASSIFIED	264909, 264595, 264683, 22279002
356	80043835 (711, 712)	Novel Protein sim. GBank gij115157 sp P16574 BVGA_BORPE - VIRULENCE FACTORS PUTATIVE POSITIVE TRANSCRIPTION REGULATOR BVGA		transcriptfactor	264909, 264591, 264592
357	80070568 (713, 714)	Novel Protein sim. GBank gij497637 (J03839) - cytochrome oxidase d subunit I [Escherichia coli]		oxidase	264605
358	37032756 (715, 716)	Novel Protein sim. GBank gij2290990 (AF006000) - Brg1 [Bordetella pertussis]		UNCLASSIFIED	264768
359	80501488 (717, 718)	Novel Protein sim. GBank gij3510639 (AF049344) - UDP- GalNAc:polypeptide N-acetylgalactosaminyltransferase T5 [Rattus norvegicus]		UNCLASSIFIED	264604, 264769
360	80026748 (719, 720)	Novel Protein sim. GBank		UNCLASSIFIED	264594
361	80584075 (721, 722)	Novel Protein sim. GBank gij113764 sp P25718 AMY1_ECOLI - ALPHA-AMYLASE PRECURSOR (1.4-ALPHA-D-GLUCAN GLUCANOHYDROLASE)		transferase	22278996, 264259, 29331822, 29331824, 264605, 55811957, 265022
362	13089485 (723, 724)	Novel Protein sim. GBank		amylase	264688
363	78750145 (725, 726)	Novel Protein sim. GBank			
364	82443593 (727, 728)	Novel Protein sim. GBank gij2829816 sp P95171 NUOK_MYCTU - NADH DEHYDROGENASE I CHAIN K (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 11) [NUO11]	Contains protein domain (PF00420) - NADH-ubiquinone/plastoquinone oxidoreductase chain 4L	dehydrogenase	264566 264769, 264602, 264604, 264508, 264762, 264638, 264488

365	88040288 (728, 730)	Novel Protein sim. GBank gij4929268[gb AAD33924.1] - (AF144237) LOMP protein [Homo sapiens]	Contains protein domain (PF00412) - LIM domain containing proteins	264488, 21906766, 21906767, 55811576, 21906769, 29148629, 22278995, 22278996, 265020, 265022, 264634, 264691, 264593, 33657023, 33657402, 264693, 264639, 264594, 29331824, 264758, 18108385, 29331827, 87168559, 265018, 22279000, 265019, 264482, 264761, 264681, 18108351, 265017, 264757
366	81821838 (731, 732)			
367	85357471 (733, 734)	Novel Protein sim. GBank gij4503843[ref NP_003908.1 pG2AD - UNKNOWN]	Contains protein domain (PF01602) - Adaptin N terminal region	60424179, 65274572, 56182575, 22278994, 56994075, 22278998, 264259, 29331822, 29331824, 56182181, 60424269, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 264905, 264828, 56182435, 265006, 264512, 265008, 264591, 55812038, 55811386, 265010, 87168559, 265017, 265018, 264604, 265019, 55811150, 264448, 264369, 264288, 264686, 264768, 56181562, 21906768, 21906769, 55811957, 35695917, 265022, 60170815, 33657023, 65274620, 18108365, 263967, 33657109, 33657349, 35695783, 264628, 18108376, 55811576, 65274791, 35695855, 56182323, 83373044, 60432113, 264563, 264564, 264567, 264509
368	78607265 (735, 736)	Novel Protein sim. GBank gij3913028[sp P94967 ALR_MYCSM - ALANINE RACEMASE	UNCLASSIFIED	264508, 264604, 264605, 264636
369	95292917 (737, 738)	Novel Protein sim. GBank gij3249559 (AF018261) - EH domain binding protein Epsin [Rattus norvegicus]		264905, 264592, 264605, 264766, 264691
370	88090966 (739, 740)	Novel Protein sim. GBank gij2995299[emb CAA18328] - (AL022268) putative tRNA delta(2)-isopentenylpyrophosphate transferase [Streptomyces coelicolor]	Contains protein domain (PF01715) - IPP transferase	264905, 264906, 264510, 264600, 264601, 264602, 264603, 265018, 264604, 264605, 265021, 264692, 264636, 264564
371	95292599 (741, 742)	Novel Protein sim. GBank gij5456934[gb AAD43716.1] - (AF152322) protocadherin gamma A2 [Homo sapiens]		264564
372	80021107 (743, 744)	Novel Protein sim. GBank gij2506393[sp P31576 FIXX_ECOLI - FERREDOXIN LIKE PROTEIN		264909
373	79863766 (745, 746)	Novel Protein sim. GBank gij3341640[emb CAA13164] - (AJ231122) z61f [Vibrio cholerae]	UNCLASSIFIED	264905, 264906
374	79847568 (747, 748)	Novel Protein sim. GBank gij5456934[gb AAD43716.1] - (AF152322) protocadherin gamma A2 [Homo sapiens]	UNCLASSIFIED	
375	91230181 (749, 750)	Novel Protein sim. GBank gij1805408[db BAA08970] - (D50453) homologues to nitrile hydratase region 3'-hypothetical protein P47K of P. chlororaphis [Bacillus subtilis]	cadherin	65274572, 264259, 29331826, 56182435, 60433356, 60433438, 264757, 55812038, 264758, 55811957, 264690, 33657023, 264769
376	80505214 (751, 752)		UNCLASSIFIED	
377	10339083 (753, 754)			264906

378	80056153 (755, 756)	Novel Protein sim. GBank gij1076013[pir][A49930 - carB (fragment)]	Contains protein domain (PF00289) - synthase	UNCLASSIFIED	265008, 264555
379	80503437 (757, 758)	Novel Protein sim. GBank gij1076013[pir][A49930 - carB (fragment)]	Contains protein domain (PF00289) - synthase	UNCLASSIFIED	264769
380	80060937 (759, 760)	Novel Protein sim. GBank gij21655[dbj][BAA02174] - (D12651) glucose dehydrogenase [Escherichia coli]	Contains protein domain (PF01011) - dehydrogenase	UNCLASSIFIED	264604
381	11769027 (761, 762)	Novel Protein sim. GBank gij3327136[dbj][BAA31636] - (AB014561) KIAA0661 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	264684
382	80054377 (763, 764)	Novel Protein sim. GBank gij3327136[dbj][BAA31636] - (AB014561) KIAA0661 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	264592
383	83259025 (765, 766)	Novel Protein sim. GBank gij3327136[dbj][BAA31636] - (AB014561) KIAA0661 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	264595, 265017, 265021, 264638, 87168518, 22279002
384	95314255 (767, 768)	Novel Protein sim. GBank gij3327136[dbj][BAA31636] - (AB014561) KIAA0661 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	264259, 29331822, 60432289, 29331827, 264288, 264768, 263987, 65274791, 35695855, 263981, 83373044, 264567
385	10237679 (769, 770)	Novel Protein sim. GBank gij1073456[pir][S47810 - probable alcohol dehydrogenase (EC 1.1.1.1) - Escherichia coli]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenases	UNCLASSIFIED	264692
386	79633434 (771, 772)	Novel Protein sim. GBank gij1073456[pir][S47810 - probable alcohol dehydrogenase (EC 1.1.1.1) - Escherichia coli]	Contains protein domain (PF00465) - iron-containing alcohol dehydrogenases	UNCLASSIFIED	264906
387	17960637 (773, 774)	Novel Protein sim. GBank gij1460074[emb][CA801049] - (Z77250) hypothetical protein Rv2566 [Mycobacterium tuberculosis]	Contains protein domain (PF01841) - Transglutaminase-like superfamily	UNCLASSIFIED	264760
388	87741376 (775, 776)	Novel Protein sim. GBank gij4240169[dbj][BAA74863.1] - (AB020647) KIAA0840 protein [Homo sapiens]	Contains protein domain (PF00846) - F-box domain.	UNCLASSIFIED	35696286, 264905, 66712502, 60432228, 264593, 60433358, 264686, 264688, 21906765, 264891, 22279000, 264482
389	78316971 (777, 778)	Novel Protein sim. GBank gij4240169[dbj][BAA74863.1] - (AB020647) KIAA0840 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	18108394, 22278996, 264630, 264556, 22279002
390	80079949 (779, 780)	Novel Protein sim. GBank gij854055[emb][CAA58337] - (X83413) U88 [Human herpesvirus 6]	UNCLASSIFIED	UNCLASSIFIED	264600
391	7657302 (781, 782)	Novel Protein sim. GBank gij854055[emb][CAA58337] - (X83413) U88 [Human herpesvirus 6]	UNCLASSIFIED	UNCLASSIFIED	264482
392	79786056 (783, 784)	Novel Protein sim. GBank gij3378523[emb][CAA08867] - (AJ009832) cyclomaltodextrinase glucanotransferase [Thermotoga neapolitana]	UNCLASSIFIED	UNCLASSIFIED	264908
393	33206031 (785, 786)	Novel Protein sim. GBank gij3378523[emb][CAA08867] - (AJ009832) cyclomaltodextrinase glucanotransferase [Thermotoga neapolitana]	UNCLASSIFIED	UNCLASSIFIED	264602, 21906764
394	10104463 (787, 788)	Novel Protein sim. GBank gij2677780 (U70327) - unknown	UNCLASSIFIED	UNCLASSIFIED	264693
395	80229010 (789, 790)	Novel Protein sim. GBank gij2677780 (U70327) - unknown	UNCLASSIFIED	UNCLASSIFIED	264508, 264563
396	20436224 (791, 792)	Novel Protein sim. GBank gij2677780 (U70327) - unknown	UNCLASSIFIED	UNCLASSIFIED	264556
397	80417014 (793, 794)	Novel Protein sim. GBank gij4507809[ref][NP_000368.1]pWASJ - Wiskott-Aldrich syndrome (eczema-thrombocytopenia)	Contains protein domain (PF00047) - struct	UNCLASSIFIED	265007, 265009, 264508, 264556, 264629, 264766
398	91230517 (795, 796)	Novel Protein sim. GBank gij1518458 (U45998) - mitochondrial solute carrier [Onchocerca volvulus]	Contains protein domain (PF00153) - transport	UNCLASSIFIED	18108398, 22278995, 22278996, 56994075, 22278999, 264259, 29331824, 29331826, 264905, 264908, 265007, 265008, 265009, 21906754, 33657084, 265017, 264448, 264288, 264766, 21906765, 21906766, 21906767, 265020, 265021, 33657023, 33657109, 264628, 35696423, 35695855, 264952, 18108380, 264567, 18108391

399	80055278 (787, 788)	Novel Protein sim. GBank gij3358091[dbj BAA31995] - (AB015974) glycerol kinase [Pseudomonas tolaasii]	Contains protein domain (PF00370) - kinase FGY family of carbohydrate kinases		264592, 264595
400	94117480 (789, 800)	Novel Protein sim. GBank gij728835[sp P39192]ALUS_HUMAN - !!! ALU SUBFAMILY SC WARNING ENTRY !!!	Contains protein domain (PF00560) - cadherin Leucine Rich Repeat		18108394, 56182575, 22278995, 22278997, 22278999, 264259, 29331824, 265006, 265007, 265009, 60432229, 33657402, 21906754, 265010, 265017, 265018, 265019, 18108351, 18108357, 21906765, 265021, 265022, 264691, 264692, 33657023, 18108370, 65274791, 264634, 264636, 60170394, 56182323, 264594
401	11397481 (801, 802)	Novel Protein sim. GBank gij4928292[gb AAD33527.1 AF13211 - (AF132117) FhuA [Staphylococcus aureus]	transport		
402	95420284 (803, 804)	Novel Protein sim. GBank gij5689487[dbj BAA83027.1] - (AB028998) KIAA1075 protein [Homo sapiens]	Contains protein domain (PF00017) - Src homology domain 2	phosphatase	65274572, 56182575, 35656286, 22278996, 22278998, 264093, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264908, 264907, 264909, 265006, 264511, 265007, 265008, 264910, 264591, 33657402, 60433356, 60433438, 264596, 21906754, 52644296, 265010, 265011, 87168559, 265017, 265018, 265019, 264681, 18108351, 264682, 264448, 264288, 264684, 264766, 264767, 264686, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 264690, 264693, 65274620, 35695763, 264628, 18108370, 264629, 18108379, 35696423, 55811576, 264635, 264636, 264557, 264639, 18108385, 22279002, 264563, 264564, 264565, 264566, 264768, 264632, 264639, 264563, 264682, 265009, 264682
403	80439913 (805, 806)		UNCLASSIFIED		
404	11809865 (807, 808)			polymerase	
405	78471280 (809, 810)	Novel Protein sim. GBank gij2661649[emb CAA15755] - (AL009198) dnaE2 [Mycobacterium tuberculosis]	Contains protein domain (PF00159) - Pancreatic hormone peptidases	UNCLASSIFIED	18108357, 264693
406	78634172 (811, 812)				
407	80478229 (813, 814)				
408	80079958 (815, 816)				
409	5640527 (817, 818)	Novel Protein sim. GBank gij3047117 (AF058919) - similar to ATP-dependent RNA helicases [Arabidopsis thaliana]		UNCLASSIFIED UNCLASSIFIED helicase	264769 264600 264259

410	95357496 (818, 820)	Novel Protein sim. GBank gij475016[jdbj BAA06184] - (D29801) Unknown [Mus musculus]	UNCLASSIFIED	264489, 52646365, 52646842, 56181686, 3596286, 52645080, 29331822, 29331824, 56182181, 29331825, 60424289, 35696052, 3356970, 264508, 264509, 264905, 264906, 264907, 264908, 52644045, 264909, 264510, 265007, 264512, 265008, 264910, 33657402, 264758, 52646317, 55811386, 265010, 265011, 265017, 264604, 265018, 55811150, 264762, 264764, 264768, 264687, 264768, 264769, 52644229, 21906768, 265020, 265021, 264534, 52644150, 264682, 33657023, 65274620, 33657109, 33657182, 27486261, 35695763, 264628, 264629, 60431528, 18108376, 263978, 35696423, 35695855, 264632, 264634, 264635, 264637, 264638, 264558, 264639, 56182323, 264559, 60432113, 22279002, 264563, 264565, 264486
411	80501670 (821, 822)		UNCLASSIFIED	264769
412	80241662 (823, 824)			264907, 264910, 263973, 22278002
413	11076446 (825, 826)	Novel Protein sim. GBank gij3261784[embj CAB08397] - (Z95558) hlpX [Mycobacterium tuberculosis]	eph	264605
414	82050554 (827, 828)	Novel Protein sim. GBank gij128038[spj P20707 DDO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	dehydrogenase	18108374, 264760, 264769, 264602, 264638, 264603, 264909, 264605
415	84453144 (829, 830)	Novel Protein sim. GBank gij4868350[gb AAD31273.1 AF13202 - (AF132025) rhophilin [Drosophila melanogaster]	UNCLASSIFIED	264908, 87168518
416	80402775 (831, 832)	Novel Protein sim. GBank gij2555172 (AF025543) - ArcC; carbamate kinase [Rhizobium etli]	kinase	264488, 264600, 264602, 264764, 264636
417	20153797 (833, 834)	Novel Protein sim. GBank gij170917[spj P52311 MTX2_XANOR - MODIFICATION METHYLASE XORII (CYTOSINE-SPECIFIC METHYLTRANSFERASE XORII) (M.XORII)]		264605
418	94125841 (835, 836)		UNCLASSIFIED	264689, 264693
419	95314273 (837, 838)		collagen	264908, 264910, 264764, 264639
420	37036349 (839, 840)	Novel Protein sim. GBank gij3261659[embj CAB03751] - (Z81388) hypothetical protein Rv2419c [Mycobacterium tuberculosis]	phosphatase	264769
421	95292942 (841, 842)	Novel Protein sim. GBank gij2916942[embj CAA17580] - (AL021999) hypothetical protein Rv0981 [Mycobacterium tuberculosis]	phosphatase	264906, 264600, 264601, 264603, 264604, 264760, 264769
422	79471293 (843, 844)	Novel Protein sim. GBank gij231752[spj Q00767 CH61_STRAL - 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL PROTEIN 1) (HSP58)]	eph	22278996, 264682, 18108376, 18108387
423	79604948 (845, 846)		UNCLASSIFIED	264509

424	78966557 (847, 848)	Novel Protein sim. GBank gil4826814[refNP_004977.1]pKTN1 - kinesin receptor			265019	
425	80431450 (849, 850)	Novel Protein sim. GBank gil1703701[jbs]178462 - KRP5-kinesin-related protein [rats, testes, Peptide Partial, 167 aa]	Contains protein domain (PF00225) - Kinesin motor domain	- struct	264909, 265007, 55811386, 264768, 55810764	
426	80064522 (851, 852)	Novel Protein sim. GBank		UNCLASSIFIED	264605, 264559	
427	80057232 (853, 854)	gil231829[sp]P29929[COBN_PSEDE - COBN PROTEIN			264603, 264636	
428	79487798 (855, 856)	Novel Protein sim. GBank gil81286[pir]S22697 - extensin - Volvox carter (fragment)		UNCLASSIFIED	264683	
429	80091252 (857, 858)	Novel Protein sim. GBank gil1806154[emb]CAB06451] - (Z84395) hypothetical protein Rv0688 [Mycobacterium tuberculosis]		UNCLASSIFIED	35696423, 35695763, 35695855, 265017, 264564, 264762	
430	80504192 (859, 860)	Novel Protein sim. GBank gil1806154[emb]CAB06451] - (Z84395) hypothetical protein Rv0688 [Mycobacterium tuberculosis]		reductase	264508, 264905, 264509, 264908, 264909, 265008, 264600, 264687, 264769, 264689, 264636, 264638, 18108385, 264488	
431	20624249 (861, 862)				264566	
432	16525372 (863, 864)				265020	
433	81494303 (865, 866)	Novel Protein sim. GBank gil3123552[emb]CAA18609] - (AL022578) dJ393P12.2 (hypothetical Proline-rich protein KIAA0269 LIKE) [Homo sapiens]		UNCLASSIFIED	264907, 264908, 264909, 264910, 264592, 264595, 264758, 264604, 264760, 264762, 264783, 264638, 264637, 22279002	
434	94326323 (867, 868)	Novel Protein sim. GBank gil2495272[sp]Q99626[CDX2_HUMAN - HOMEBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEBOX PROTEIN 2) (CDX-3)]	Contains protein domain (PF00169) - PH domain	UNCLASSIFIED	55812038, 56182181, 56181562, 29331828, 35696052, 55810764, 55811576, 65274791, 35695855, 60432113, 55811150, 264636, 264766	
435	80502738 (869, 870)	Novel Protein sim. GBank gil114105[sp]P08532[ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		transport	264595, 264769	
436	41085953 (871, 872)			UNCLASSIFIED	265020, 22279002	
437	11399291 (873, 874)			UNCLASSIFIED	264593	
438	11773835 (875, 876)			UNCLASSIFIED	264686	
439	80018495 (877, 878)	Novel Protein sim. GBank gil3242702 (AC003040) - hypothetical protein [Arabidopsis thaliana]			264905, 264600, 264602, 264604	
440	79841062 (879, 880)	Novel Protein sim. GBank gil2291232[gb]AB55351.1] - (AF016427) Contains similarity to Pfam domain: PF00004 (AAA). Score=268.1, E-value=3.7e-77, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00004) - ATPases associated with various cellular activities (AAA)	ATPase-associated	35696052, 264905, 264908, 264909, 265011, 35696423	
441	20396835 (881, 882)	Novel Protein sim. GBank gil5639946[gb]AAD45904.1[AF16132 - (AF161328) histidine kinase CstS [Corynebacterium diphtheriae]			264605	
442	85281058 (883, 884)	Novel Protein sim. GBank gil1184790 (U46068) - von Ebner minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909	
443	82456427 (885, 886)	Novel Protein sim. GBank gil5689993[emb]CAB52056.1] - (AL109732) putative ATP-binding RNA helicase [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	35696052, 264508, 264906, 264512, 264604, 264762, 264769, 264689, 264636	
444	11395897 (887, 888)	Novel Protein sim. GBank gil1783249[db]BAA11726] - (D83026) homologous to citrate-sodium symport (citrate transporters); hypothetical [Bacillus subtilis]		UNCLASSIFIED	264591	

445	7952708 (889, 890)	Novel Protein sim. GBank gi 5531272 emb CAB50897.1  - (AJ243800) WSCA homologue [Kluyveromyces fragilis]		UNCLASSIFIED	264693
446	79810937 (891, 892)	Novel Protein sim. GBank gi 538413 (L36315) - zinc finger protein [Mus musculus]	Contains protein domain (PF000096) - Zinc finger, C2H2 type		264509
447	80438888 (893, 894)	Novel Protein sim. GBank gi 538413 (L36315) - zinc finger protein [Mus musculus]		transcript factor	264768, 55811576
448	80238110 (895, 896)	Novel Protein sim. GBank gi 1542914 emb CAB02185  - (Z80108) fnt [Mycobacterium tuberculosis]	Contains protein domain (PF00551) - Formyl transferase	dehydrogenase	264508, 264600, 264603, 264605, 264682, 264768, 18108362, 264634, 18108387
449	20460634 (897, 898)	Novel Protein sim. GBank gi 118794 sp P10443 DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	264605, 264559
450	94631210 (899, 900)	Novel Protein sim. GBank gi 4589508 dbj BAA76775.1  - (AB023148) KIAA0931 protein [Homo sapiens]	Contains protein domain (PF00481) - Protein phosphatase 2C	phosphatase	65274572, 22278998, 29331824, 29331826, 264906, 264910, 264592, 52646317, 265017, 21906767, 55811957, 56526486, 22279002
451	21433609 (901, 902)			UNCLASSIFIED	264486
452	10267276 (903, 904)				264692
453	52560096 (905, 906)	Novel Protein sim. GBank gi 2650614 (AE001104) - conserved hypothetical protein [Archaeoglobus fulgidus]		UNCLASSIFIED	264907, 264600
454	39523922 (907, 908)	Novel Protein sim. GBank gi 2493000 sp Q09450 SCOT_CAEEL - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (3-OXOACID COA-TRANSFERASE)	transferase		264603
455	13089692 (909, 910)			UNCLASSIFIED	264687
456	79563081 (911, 912)			UNCLASSIFIED	264691
457	79831273 (913, 914)	Novel Protein sim. GBank gi 4468699 emb CAB38153.1  - (AL035591) putative integral membrane export protein [Streptomyces coelicolor]			264905
458	79581227 (915, 916)	Novel Protein sim. GBank gi 3411053 (AF034863) - synaptic scaffolding molecule [Rattus norvegicus]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF)	kinase	55812038, 265010, 265018, 264681
459	80567359 (917, 918)	Novel Protein sim. GBank gi 4506075 ref NP_002733.1 pPRKC - protein kinase C, mu	Contains protein domain (PF00130) - Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	22278997, 264259, 29331826, 265018, 264448, 264369, 21806765, 35696423
460	79245690 (919, 920)	Novel Protein sim. GBank gi 113158 sp P25516 ACO1_ECOLI - ACONITATE HYDRATASE 1 (CITRATE HYDRO-LYASE 1) (ACONITASE 1)		UNCLASSIFIED	264906
461	95287618 (921, 922)	Novel Protein sim. GBank gi 1168574 sp P42464 ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	264602, 264605, 264768, 264769, 265021, 33657023, 264559

462	7960589 (923, 924)	Novel Protein sim. GBank gij1346891sp P45597 PTF1_XANCP - MULTIPHOSPHORYL TRANSFER PROTEIN (MTP) (CONTAINS: PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I); PHOSPHOCARRIER PROTEIN HPR (PROTEIN H); PTS SYSTEM, FRUCTOSE-SPECIFIC IIA COMPONENT ...	Contains protein domain (PF00391) - PEP-utilizing enzymes	UNCLASSIFIED	264807
463	79786417 (925, 926)	Novel Protein sim. GBank gij183413 U88 [Human herpesvirus 6]		UNCLASSIFIED	264905, 264906, 264908, 264909, 264910, 264591, 264595, 265011, 264632, 264635, 264636, 264637, 264638, 264639
464	82340151 (927, 928)	Novel Protein sim. GBank gij568976 emb CAA58337  -		UNCLASSIFIED	264634
465	83005730 (928, 930)	(AJ242832) calpain [Homo sapiens]	Contains protein domain (PF00848) - Calpain family cysteine protease		265017, 21906764, 265020
466	20460645 (931, 932)	Novel Protein sim. GBank gij1808175 emb CAA06470  - (Z84395) rpsC [Mycobacterium tuberculosis]	Contains protein domain (PF00417) - Ribosomal protein S3, N-terminal domain.		264605, 264559
467	80408035 (933, 934)	Novel Protein sim. GBank gij548705 sp P36949 IRBSB_BACSU - D-RIBOSE-BINDING PROTEIN PRECURSOR		UNCLASSIFIED	264764
468	52562208 (935, 936)	Novel Protein sim. GBank gij2114024 emb CAA08957  - (Z95558) grcC1 [Mycobacterium tuberculosis]		UNCLASSIFIED	264692
469	19520527 (937, 938)	Novel Protein sim. GBank gij2909459 emb CAA17347  - (AL021929) cabQ [Mycobacterium tuberculosis]		UNCLASSIFIED	264488
470	80502756 (939, 940)	Novel Protein sim. GBank gij114921sp P17447 BETT_ECOLI - HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN	synthase		264602, 264769
471	17637351 (941, 942)		transport		265019
472	80047458 (943, 944)	Novel Protein sim. GBank gij862343 (L10908) - Gcap1 gene product [Mus musculus]		UNCLASSIFIED	264596, 264685, 264557
473	20558783 (945, 946)				264369
474	80593365 (947, 948)			UNCLASSIFIED	22278997, 264692, 264288
475	82454665 (949, 950)			UNCLASSIFIED	264907, 264908, 264511, 265009, 264762, 264448, 264636, 264638
476	94143857 (951, 952)	Novel Protein sim. GBank gij5453656 ref NP_006329.1 pGAC1 - glioma amplified on chromosome 1 protein (leucine-rich)	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	65274572, 60432049, 264259, 264508, 52644045, 55812038, 264758, 265011, 264288, 264686, 52844229, 85274791, 264638, 264566
477	79175833 (953, 954)			UNCLASSIFIED	264636
478	79633483 (955, 956)			UNCLASSIFIED	264690, 264693
479	80189746 (957, 958)			collagen	264686, 35695855, 265008, 264631, 264910, 264632, 264638, 265018, 264369, 264909
480	79390729 (959, 960)	Novel Protein sim. GBank gij1127551 (U18939) - orf2 [Batrachocottus baikalensis]		mapolymerase	264369
481	79624578 (961, 962)			UNCLASSIFIED	264693
482	83050611 (963, 964)	Novel Protein sim. GBank gij4063042 (AF068065) - GP900; mucin-like glycoprotein [Cryptosporidium parvum]		UNCLASSIFIED	264909, 264686, 264768, 264693, 55811576, 56182323, 18108385



483	20293306 (965, 966)	Novel Protein sim. GBank gjl2104303[emb]CAB08632] - (Z95387) hypothetical protein Rv2610c [Mycobacterium tuberculosis]	Contains protein domain (PF00534) - Glycosyl transferases group 1	264600	
484	11618046 (967, 968)	Novel Protein sim. GBank gjl3450883 (AF083334) - fibroin [Antheraea pernyi]		UNCLASSIFIED	264594
485	80191234 (969, 970)	Novel Protein sim. GBank gjl5042272[emb]CAB44526.1] - (AL078616) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]		UNCLASSIFIED	264369, 21906765, 22279000, 22279002
486	80059042 (971, 972)			dehydrogenase	264604
487	11813339 (973, 974)				264638
488	91222383 (975, 976)	Novel Protein sim. GBank gjl5724778[gb]AAC53522.2] - (AF012273) rho-type GTPase-activating protein rhoGAPX-1 [Mus musculus]	Contains protein domain (PF00620) - RhoGAP domain	264686, 66714117, 264768, 18108385, 55811576, 265006, 265008, 265009, 265019, 22279002, 264259, 18108370, 264907, 264764, 56182323, 264288, 264693	
489	10867710 (977, 978)	Novel Protein sim. GBank gjl3882223[db]EAA34471.1] - (AB018294) KIAA0751 protein [Homo sapiens]		kinase	264639
490	95361124 (979, 980)	Novel Protein sim. GBank gjl82091[pir]A25494 - hydroxyproline-rich glycoprotein - tomato (fragment)		collagen	22278996, 29331822, 29331828, 264107, 264909, 264110, 265009, 264592, 264593, 60433358, 264288, 264693, 263974, 263976, 20281071, 60432113
491	80498412 (981, 982)	Novel Protein sim. GBank gjl2894206[emb]CAA17072] - (AL021840) hypothetical protein Rv3258c [Mycobacterium tuberculosis]		UNCLASSIFIED	264769
492	87421264 (983, 984)				264600
493	11692942 (985, 986)			UNCLASSIFIED	264638
494	87726604 (987, 988)	Novel Protein sim. GBank gjl5262605[emb]CAB45743.1] - (AL080150) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264489, 35696286, 60432289, 29331828, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 33657402, 264762, 264764, 264768, 264769, 264688, 21906765, 21906769, 35695917, 265020, 264693, 33657109, 264629, 35695423, 35695855, 264634, 264638
495	80028599 (989, 990)	Novel Protein sim. GBank gjl2791517[emb]CAA16054] - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	Contains protein domain (PF00005) - ABC transporter	transport	264602, 264682, 264638
496	7895624 (991, 992)	Novel Protein sim. GBank gjl230281[pdb]1R69] - 434 Repressor (Amino-Terminal Domain) (R1-69)	Contains protein domain (PF01381) - Helix-turn-helix		264601, 265021
497	78949661 (993, 994)	Novel Protein sim. GBank gjl128736[sp]P28225[PDXH_ECOLI - PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (PNP/PMP OXIDASE)]		oxidase	265006

498	88095488 (895, 896)	Novel Protein sim. GBank gij1145789 (U41662) - neuroligin 2 [Rattus norvegicus]	Contains protein domain (PF00135) - Carboxylesterases	esterase	264259, 29331826, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 264591, 33657402, 264758, 265010, 265011, 264600, 264601, 264605, 264683, 264764, 264766, 264767, 264768, 264687, 264769, 21906767, 33657023, 284693, 264628, 284629, 35696423, 264630, 264632, 264634, 284635, 264637, 264638, 264558, 264639, 18108385, 264563, 284564, 264565, 264566, 264567
499	20438222 (987, 988)	Novel Protein sim. GBank gij97480[pil]S19739 - integral membrane protein - Rhodobacter capsulatus		UNCLASSIFIED	264605
500	11076810 (999, 1000)				264605
501	13418034 (1001, 1002)	Novel Protein sim. GBank gij5708250[emb]CAB52363.1] - (AL108747) putative integral membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264688
502	80021176 (1003, 1004)	Novel Protein sim. GBank gij4468678[emb]CAB38132.1] - (AL035591) glucose-6-phosphate isomerase [Streptomyces coelicolor]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	22278996, 265011, 264602, 264605, 264635
503	20264483 (1005, 1006)			UNCLASSIFIED	264564
504	10887321 (1007, 1008)			UNCLASSIFIED	264687
505	95003068 (1009, 1010)			UNCLASSIFIED	264369
506	16454292 (1011, 1012)			UNCLASSIFIED	265010
507	20451598 (1013, 1014)	Novel Protein sim. GBank gij4033509[sp]P02598[CALM_TETPY - CALMODULIN gij2501069[sp]Q46127[SYW_CLOLO - TRYPTOPHANYL-TRNA SYNTHETASE (TRYPTOPHAN--TRNA LIGASE) (TRPRS)]	Contains protein domain (PF00036) - EF hand	UNCLASSIFIED	264604
508	79841424 (1015, 1016)	Novel Protein sim. GBank gij466068[sp]P34618[YO82 CAEEL - HYPOTHETICAL 33.8 KD PROTEIN ZK1235.2 IN CHROMOSOME III]		UNCLASSIFIED	264908
509	11776386 (1017, 1018)				264638
510	83373465 (1019, 1020)			UNCLASSIFIED	264687, 264639
511	16525578 (1021, 1022)				265007
512	20399484 (1023, 1024)	Novel Protein sim. GBank gij2497419[sp]P55635[Y4RB_RHISN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB]		UNCLASSIFIED	264565
513	79457404 (1025, 1026)	Novel Protein sim. GBank gij1276897 (U41809) - cyclin J [Drosophila melanogaster]	Contains protein domain (PF00134) - Cyclin	cyclin	264683, 264689, 35696423, 264639
514	79813805 (1027, 1028)	Novel Protein sim. GBank gij1184780 (U46068) - von Ebner minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909
515	79462591 (1029, 1030)				22278999, 264690
516	9862020 (1031, 1032)	Novel Protein sim. GBank gij2127400[pil]S65770 - maltodigosyltrehalose trehalohydrolase - Arthrobacter sp. (strain Q36)		amylase	264910

517	95292994 (1033, 1034)	Novel Protein sim. GBank gi 29938605 (AE000725) - ribose 5 phosphate isomerase B [Aquifex aeolicus]		isomerase	265018, 264605, 264764, 264766, 264687, 264891, 264585
518	8491831 (1035, 1036)	Novel Protein sim. GBank gi 854065 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264487
519	91677888 (1037, 1038)	Novel Protein sim. GBank gi 5689365 dbj BAA83073.1  - (AB024075) B120 [Homo sapiens]	Contains protein domain (PF01388) - ARID DNA binding domain	dna_rna_bind	52644507, 22278997, 22278998, 60432049, 264258, 52645080, 29331824, 66714117, 60424269, 29331826, 35696052, 264905, 29331830, 66712502, 264511, 265007, 264591, 60432228, 33657402, 60433438, 21908754, 33109854, 52644296, 67168474, 87168559, 285017, 265018, 264604, 265019, 264681, 264448, 264369, 264288, 264685, 21906765, 21906766, 21906767, 21906769, 265021, 60170615, 33657023, 264692, 52645129, 33657109, 27486262, 27486264, 35695763, 18108370, 264629, 52644332, 56182323, 264639, 83373044, 18108385, 56526486, 60432113
520	79889188 (1039, 1040)	Novel Protein sim. GBank gi 1169126 sp P48839 CTPA_MYCLE - CATION-TRANSPORTING P-TYPE ATPASE A		transport	264769
521	11076821 (1041, 1042)	Novel Protein sim. GBank gi 1172869 sp P44433 RBKS_HAEIN - RIBOKINASE	Contains protein domain (PF00294) - pfkB family carbohydrate kinase	kinase	264805, 264768
522	80435060 (1043, 1044)	Novel Protein sim. GBank gi 1172869 sp P44433 RBKS_HAEIN - RIBOKINASE		UNCLASSIFIED	264829
523	18356013 (1045, 1046)	Novel Protein sim. GBank gi 2132243 pir J561028 - hypothetical protein YPL236c - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED	
524	80261805 (1047, 1048)	Novel Protein sim. GBank gi 4033608 dbj BAA35136  - (AB012308) B2HC [Anthracinus crassispina]		ATPase_associated	264092, 264596, 265011
525	78810046 (1049, 1050)	Novel Protein sim. GBank gi 4106610 emb CAA21365  - (AL031866) ORF42, len=386 aa, similarity to an aminotransferase, in P95957 Sulfolobus solfataricus, (401 aa), 33.1% identity in 393 aa overlap. FastA scores: opt:468, E(): 8.5e-24, in O64602 R. norvegicus (425 aa), 28.6% identl...		UNCLASSIFIED	264907
526	36827630 (1051, 1052)	Novel Protein sim. GBank gi 4106610 emb CAA21365  - (AL031866) ORF42, len=386 aa, similarity to an aminotransferase, in P95957 Sulfolobus solfataricus, (401 aa), 33.1% identity in 393 aa overlap. FastA scores: opt:468, E(): 8.5e-24, in O64602 R. norvegicus (425 aa), 28.6% identl...		UNCLASSIFIED	264758
527	80504728 (1053, 1054)	Novel Protein sim. GBank gi 731088 sp P24215 UXUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		UNCLASSIFIED	264769
528	85484134 (1055, 1056)	Novel Protein sim. GBank gi 731088 sp P24215 UXUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		UNCLASSIFIED	56182575, 265017, 265018
529	17936810 (1057, 1058)	Novel Protein sim. GBank gi 731088 sp P24215 UXUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		hydrolase	265019
530	10887336 (1059, 1060)	Novel Protein sim. GBank gi 42144 emb CAA25200  - (X00513) NusA protein (nusA) [Escherichia coli]		UNCLASSIFIED	264687
531	80226576 (1061, 1062)	Novel Protein sim. GBank gi 5262640 emb CAB45758.1  - (AL080170) hypothetical protein [Homo sapiens]	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	264555, 264556, 264557, 264558, 18108385
532	90933444 (1063, 1064)	Novel Protein sim. GBank gi 5262640 emb CAB45758.1  - (AL080170) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264488, 264490, 264259, 264592, 264760, 265021, 264690, 263976, 264558

533	87761531 (1065, 1066)	Novel Protein sim. GBank gi 4883636 gb AAD31593.1 AF11229 - (AF11229) integral inner nuclear membrane protein MAN1 [Homo sapiens]				264907, 264909, 264768, 35695917, 264630, 264555
534	82368264 (1067, 1068)	Novel Protein sim. GBank gi 2995352 emb CAA04606.1  - (AJ001206) pep1 [Streptomyces coelicolor]			UNCLASSIFIED	264905, 265011, 264601, 264602, 264605, 264762, 264768, 265020, 264693, 264636 264906
535	79641850 (1069, 1070)	Novel Protein sim. GBank gi 3878638 emb CAA88953  - (Z49128) similar to cAMP-dependent protein kinase; cDNA EST EMBL: T00719 comes from this gene; cDNA EST yk465d8.3 comes from this gene; cDNA EST yk465d8.5 comes from this gene; cDNA EST yk49214.3 comes from this gene; cDNA EST y...		Contains protein domain (PF00069) - ATPase associated Eukaryotic protein kinase domain		
536	79907207 (1071, 1072)	Novel Protein sim. GBank gi 2495628 sp P55757 YOH1_SERMA - HYPOTHETICAL 10.1 KD PROTEIN IN BIOA 5 REGION			reductase	18108376, 264905, 264906, 264907, 264909
537	94147448 (1073, 1074)					265008, 264605, 65274781
538	87821963 (1075, 1076)	Novel Protein sim. GBank gi 134920 sp P21997 SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)		Contains protein domain (PF00595) - POZ domain (Also known as DHR or GLGF).	collagen	29331822, 29331824, 29331825, 29331826, 29331827, 264908, 52644045, 33657402, 265017, 264762, 264683, 264288, 264685, 21906765, 35695763, 264558, 60170394, 264559, 22279002 264602, 265019
539	28366269 (1077, 1078)	Novel Protein sim. GBank gi 2498433 sp Q12341 HAT1_YEAST - HISTONE ACETYLTRANSFERASE			histone	
540	79637077 (1079, 1080)					264693
541	87762268 (1081, 1082)	Novel Protein sim. GBank gi 398224 dbj BAA34480.1  - (AB018303) KIAA0760 protein [Homo sapiens]		Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	18108394, 22278997, 22278998, 264259, 264112, 265009, 33657402, 55812038, 52646317, 265017, 21906765, 264693, 55811576, 264635, 56526486, 264568 264910, 265018, 264689, 264638, 264486
542	95295836 (1083, 1084)	Novel Protein sim. GBank gi 5042272 emb CAB44526.1  - (AL078618) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]			dehydrogenase	
543	78786280 (1085, 1086)					
544	20437191 (1087, 1088)	Novel Protein sim. GBank gi 279139 emb CAA15994  - (AL021184) hypothetical protein Rv1464 [Mycobacterium tuberculosis]			UNCLASSIFIED	264602, 264908
545	80434504 (1089, 1090)				UNCLASSIFIED	264605
546	80249016 (1091, 1092)	Novel Protein sim. GBank gi 488721 gb AAD32237.1 AF14744 - (AF14744) penicillin binding protein 1B [Pseudomonas aeruginosa]				264768, 264634, 264907, 264592, 264909 264600, 264602, 21806765
547	11077563 (1093, 1094)	Novel Protein sim. GBank gi 1350855 sp P19176 IRPOC_PSEPU - DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (TRANSCRIPTASE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT)			mapolymerase	264604
548	82114936 (1095, 1096)	Novel Protein sim. GBank gi 2330021 (AF019250) - kinesin- related protein; KRP. Costal2 [Drosophila melanogaster]			UNCLASSIFIED	264488, 264905, 264910, 264760, 264693, 264639, 264563, 264564

549	95421904 (1097, 1098)	Novel Protein sim. GBank gi 4337460 gb AAD18133  - (AF056195) neuroblastoma-amplified protein [Homo sapiens]		UNCLASSIFIED	264488, 65274572, 18108398, 22278995, 22278996, 22278997, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 29331826, 35696052, 265007, 265008, 264910, 264592, 33657402, 33109954, 265017, 265018, 265019, 18108351, 264448, 264764, 264369, 264288, 264766, 264686, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 264691, 33657023, 264692, 264693, 65274620, 52845129, 33657109, 27486261, 27486262, 27486264, 33657349, 55811576, 18108387, 60432113, 22279002
550	10866616 (1099, 1100)	Novel Protein sim. GBank		UNCLASSIFIED	264688
551	60439990 (1101, 1102)	Novel Protein sim. GBank gi 3122893 sp P94985 SYFB_MYCTU - PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (PHENYLALANINE--TRNA LIGASE BETA CHAIN) (PHERS)		UNCLASSIFIED	264908, 264909, 264768
552	94672870 (1103, 1104)	Novel Protein sim. GBank gi 552087 (M33753) - crumbs protein [Drosophila melanogaster]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264689, 264639, 264563
553	80106002 (1105, 1106)	Novel Protein sim. GBank gi 5019771 gb AAD37857.1 AF13326 - (AF133263) histidine protein kinase-response regulator hybrid protein CvgSY [Pseudomonas syringae pv. syringae]	kinase		55811957, 264628
554	79818379 (1107, 1108)	Novel Protein sim. GBank gi 131515 sp P02908 PTGA_SALTY - PTS SYSTEM, GLUCOSE-SPECIFIC IIA COMPONENT (EIIA-GLC) (GLUCOSE-PERMEASE IIA COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, A COMPONENT) (EIIA-GLC)			264906
555	78966347 (1109, 1110)	Novel Protein sim. GBank gi 3914014 sp P96380 MFD_MYCTU - TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)	Contains protein domain (PF00358) - phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 1	transport	264762
556	20457127 (1111, 1112)	Novel Protein sim. GBank gi 170933 sp P45331 METE_HAEIN - 5-METHYL-TETRAHYDROPTEROYL-TRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME) (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE)		transcriptfactor	284508, 264605, 284559
557	19523405 (1113, 1114)	Novel Protein sim. GBank gi 5042273 emb CAB44527.1  - (AL078518) nuoE, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	264488
558	20724429 (1115, 1116)	Novel Protein sim. GBank gi 4980567 gb AAD35173.1 AE00169 - (AE001694) iron(ii) ABC transporter, permease protein [Thermotoga maritima]		UNCLASSIFIED	264602
559	80084353 (1117, 1118)	Novel Protein sim. GBank gi 4980567 gb AAD35173.1 AE00169 - (AE001694) iron(ii) ABC transporter, permease protein [Thermotoga maritima]		UNCLASSIFIED	264634

560	80066533 (1119, 1120)	Novel Protein sim. GBank gi 2492595 sp Q53183 Y4TR_RHISN - PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y4TR	Contains protein domain (PF000005) - transport ABC transporter	18108396, 264806, 264602, 264604, 18108374
561	20293187 (1121, 1122)		UNCLASSIFIED	264600
562	11698161 (1123, 1124)		UNCLASSIFIED	264689
563	79761420 (1125, 1126)	Novel Protein sim. GBank gi 4104925 (AF042276) - poly(hydroxycarboxylate) granule associated protein GA2 [Pseudomonas putida]	UNCLASSIFIED	264910, 264691
564	56716390 (1127, 1128)	Novel Protein sim. GBank gi 2792310 (AF040570) - unknown [Amycolatopsis mediterranei]	dehydrogenase	264592
565	56465616 (1129, 1130)	Novel Protein sim. GBank gi 3448294 db BAA32462  - (AB011532) MEGF8 [Rattus norvegicus]	synthase	265010
566	94323888 (1131, 1132)	Novel Protein sim. GBank gi 4539588 emb CAB38487.1  - (AL035636) putative helicase [Streptomyces coelicolor]	helicase	264909, 264510, 265008, 264910, 264758, 264600, 264602, 264604, 264605, 264768, 264687, 264689, 35695917, 264693, 65274620, 264486
567	79560955 (1133, 1134)	Novel Protein sim. GBank gi 100506 pir S17455 - Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) (EC 1.1.1.40) - Flavaria trinevia (fragment)	UNCLASSIFIED	264681, 264691, 264593
568	94681793 (1135, 1136)	Novel Protein sim. GBank gi 3915843 sp O31212 RS2_STRCO - 30S RIBOSOMAL PROTEIN S2	Contains protein domain (PF00390) - Malic enzyme	264689
569	39508897 (1137, 1138)	Novel Protein sim. GBank gi 115122 sp P21627 BRAD_PSEAE - HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN BRAD	Contains protein domain (PF00318) - Ribosomal protein S2	264565
570	78375927 (1139, 1140)	Novel Protein sim. GBank	UNCLASSIFIED	18108376, 18108387, 264565
571	79793981 (1141, 1142)		transport	264907, 264909
572	36996838 (1143, 1144)		UNCLASSIFIED	264762
573	20715521 (1145, 1146)	Novel Protein sim. GBank gi 4539223 emb CAB39881.1  - (AL049497) putative integral membrane protein [Streptomyces coelicolor]	UNCLASSIFIED	265007, 264601
574	13521592 (1147, 1148)			264636
575	13076416 (1149, 1150)	Novel Protein sim. GBank gi 118794 sp P10443 DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN	polymerase	264687
576	20482246 (1151, 1152)	Novel Protein sim. GBank gi 5457625 emb CAB49116.1  - (AJ248283) PAB2227 [Pyrococcus abyssi]		264605
577	56727102 (1153, 1154)	Novel Protein sim. GBank gi 5042274 emb CAB44528.1  - (AL078618) nuod, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00346) - Respiratory-chain NADH dehydrogenase, 49 Kd subunit	35696052, 264636
578	11804477 (1155, 1156)			264638
579	11794723 (1157, 1158)	Novel Protein sim. GBank gi 1723081 sp Q11046 Y089_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.09	transport	264682, 264556

580	80059417 (1159, 1160)						22278999, 35696052, 264555, 264556, 264558
581	79230833 (1161, 1162)					UNCLASSIFIED	265008, 264564
582	80049617 (1163, 1164)	Novel Protein sim. GBank gij3243131 (AF045777) - ilin [Drosophila melanogaster]	Contains protein domain (PF00047) - struct			struct	265021, 264555, 264557
583	79321392 (1165, 1166)	Novel Protein sim. GBank gij2501182[sp]P7726[YAJR_ECOLI - HYPOTHETICAL 49.0 KD PROTEIN IN ABPA-CYOE INTERGENIC REGION]	transport			transport	264594
584	79845024 (1167, 1168)						
585	79581454 (1169, 1170)	Novel Protein sim. GBank gij388222[jdb]BAA34470.1]- (AB018293) KIAA0750 protein [Homo sapiens]	UNCLASSIFIED			UNCLASSIFIED	264488, 264908, 264768, 264887, 35696423
586	38277486 (1171, 1172)					UNCLASSIFIED	265018, 264684, 21808769
587	80497359 (1173, 1174)	Novel Protein sim. GBank gij4467250[emb]CAB37575]- (AL035569) probable Glu-RNA Gln amidotransferase subunit [Streptomyces coelicolor]	hydrolase			hydrolase	264908, 265007
588	79557239 (1175, 1176)	Novel Protein sim. GBank gij5689519[jdb]BAA83043.1]- (AB028014) KIAA1091 protein [Homo sapiens]	UNCLASSIFIED			UNCLASSIFIED	264600, 264602, 264605, 264769, 264690, 264557
589	79805828 (1177, 1178)					UNCLASSIFIED	265020, 264692
590	79815629 (1179, 1180)					UNCLASSIFIED	22278996, 264907, 264909, 264510, 265009, 265010, 264687, 264769, 35695917, 18108376, 264634, 264638, 264638
591	10313540 (1181, 1182)	Novel Protein sim. GBank gij2143293[emb]CAB09390]- (Z95972) rpoB [Mycobacterium tuberculosis]	mapolymerase			UNCLASSIFIED	264906, 264909
592	13889767 (1183, 1184)					MHC	263972
593	82348698 (1185, 1186)	Novel Protein sim. GBank gij4511983[jdb]A021543.1]- (AF088896) electrotransfer ubiquinone oxidoreductase [Zymomonas mobilis]	dehydrogenase			dehydrogenase	264511, 264762, 264769, 264486
594	20212392 (1187, 1188)	Novel Protein sim. GBank gij1272368 (U51896) - LfGE [Vibrio parahaemolyticus]	UNCLASSIFIED			UNCLASSIFIED	264605
595	10064064 (1189, 1190)	Novel Protein sim. GBank gij131490[sp]P20966[PTFB_ECOLI - PTS SYSTEM, FRUCTOSE-SPECIFIC IIBC COMPONENT (EIIBC-FRU) (FRUCTOSE-PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EII-FRU)]					264769
596	13085170 (1191, 1192)					UNCLASSIFIED	264636
597	80259003 (1193, 1194)					UNCLASSIFIED	264592
598	94140216 (1195, 1196)					UNCLASSIFIED	264758, 55810764, 264555, 264556, 264637, 83373044
599	20385137 (1197, 1198)	Novel Protein sim. GBank gij125329[sp]P04951[KDSB_ECOLI - 3-DEOXY-MANNO- OCTULONATE CYTIDYLTRANSFERASE (CMP-KOO SYNTHETASE) (CMP-2-KETO-3-DEOXYOCTULONIC ACID SYNTHETASE) (CKS)]	UNCLASSIFIED			UNCLASSIFIED	264603
600	10357663 (1199, 1200)						
601	79610404 (1201, 1202)	Novel Protein sim. GBank gij2127414[pir]S60064 - hypothetical protein 2 - Corynebacterium glutamicum	UNCLASSIFIED			UNCLASSIFIED	264906, 264510

602	79250602 (1203, 1204)	Novel Protein sim. GBank gjl3522861gb AAC34243.1  - (AC004411) putative pto kinase [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	265007
603	11466067 (1205, 1206)			UNCLASSIFIED	264595
604	81675420 (1207, 1208)			UNCLASSIFIED	264758
605	20436657 (1209, 1210)	Novel Protein sim. GBank gjl1175322sp P44917 Y883_HAEIN - HYPOTHETICAL PROTEIN HI0883		UNCLASSIFIED	264605
606	80334582 (1211, 1212)	Novel Protein sim. GBank gjl5020284gb AAD38043.1 AF15136 - (AF151363) Cdc42 GTPase-activating protein [Mus musculus]		UNCLASSIFIED	264764
607	85361506 (1213, 1214)	Novel Protein sim. GBank gjl188864 (M74027) - mucin [Homo sapiens]		UNCLASSIFIED	264508, 264906, 85658542, 264682, 264687, 264689, 264534, 18108376, 35696423, 264636, 264555, 264638
608	11810888 (1215, 1216)			UNCLASSIFIED	264682
609	80084775 (1217, 1218)	Novel Protein sim. GBank gjl2486701sp P55552 Y4LL_RHISN - HYPOTHETICAL 91.8 KD PROTEIN Y4LL	Contains protein domain (PF00989) - PAS domain	UNCLASSIFIED	264605
610	79629413 (1219, 1220)				264692
611	87586205 (1221, 1222)				264508, 264905, 264907, 264908, 264909, 264511, 264910, 264758, 264604, 264684, 264766, 264689, 264692, 264628, 264635, 264636, 264637, 264558
612	85287851 (1223, 1224)	Novel Protein sim. GBank gjl1877366 emb CAB07118  - (Z92772) recD [Mycobacterium tuberculosis]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	nuclease	264600, 264601, 264604, 264769, 264558, 264565
613	7523475 (1225, 1226)			UNCLASSIFIED	264369
614	78969348 (1227, 1228)	Novel Protein sim. GBank gjl5114231gb AAD40238.1 AF13670 - (AF136709) histidine kinase YycG [Staphylococcus aureus]		kinase	18108372, 264563
615	39586896 (1229, 1230)	Novel Protein sim. GBank gjl1339950 dbj BAA12741  - (D85230) large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]		synthase	264600, 264602, 264629
616	20465331 (1231, 1232)	Novel Protein sim. GBank gjl544367sp P35673 GALE_ERWAM - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)		isomerase	264605
617	91227222 (1233, 1234)	Novel Protein sim. GBank gjl2488097sp Q60769 TNP3_MOUSE - TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 3 (PUTATIVE DNA BINDING PROTEIN A20) (ZINC FINGER PROTEIN A20)	Contains protein domain (PF00641) - Int Zn-finger in Ran binding protein and others.	Int	52845156, 21906765, 35696423, 21906768, 21906769, 22278994, 35696286, 22278996, 265020, 265021, 265007, 265008, 264636, 52844150, 33657023, 264692, 264693, 29331822, 29331824, 55812038, 83373044, 56182181, 60424269, 66714117, 29331825, 33657109, 29331826, 33657182, 29331827, 35696052, 29331828, 27486262, 33657349, 56526486, 265018, 265019, 22279002, 264482, 264448, 29331830, 66712502, 264909



618	20632843 (1235, 1236)	Novel Protein sim. GBank gij5459388[emb]CAB50746.1] - (AL096839) putative aminotransferase [Streptomyces coelicolor]		isomerase	264603	
619	91227224 (1237, 1238)				56994075, 29331826, 33656970, 265008, 33657402, 33109954, 87108559, 264448, 18108374, 83373044	
620	81183143 (1239, 1240)	Novel Protein sim. GBank gij464335[sp]Q05922[DUS2_MOUSE - DUAL SPECIFICITY PROTEIN PHOSPHATASE 2 (DUAL SPECIFICITY PROTEIN PHOSPHATASE PAC-1)]		phosphatase	29146498, 264758, 264368, 29148627	
621	80239251 (1241, 1242)			UNCLASSIFIED	264556, 264558, 264639	
622	20456427 (1243, 1244)	Novel Protein sim. GBank gij2633557[emb]CAB13060] - (Z99110) ydf [Bacillus subtilis]		UNCLASSIFIED	264605	
623	10131788 (1245, 1246)	Novel Protein sim. GBank gij1857710[gb]AAB48482] - (U87224) contactin associated protein [Rattus norvegicus]	Contains protein domain (PF00054) - Laminin G domain	laminin	264906	
624	19534127 (1247, 1248)	Novel Protein sim. GBank gij1705703[sp]P52225[CCMF_PSEFL - CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK]		cytochrome	264596	
625	13084619 (1249, 1250)	Novel Protein sim. GBank gij2894252[emb]CAA17114.1] - (AL021841) hypothetical protein Rv3342 [Mycobacterium tuberculosis]		UNCLASSIFIED	264688	
626	88062603 (1251, 1252)	Novel Protein sim. GBank gij416592[sp]P32323JAGA1_YEAST - A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR		UNCLASSIFIED	29331822, 264905, 264908, 33657023, 33657109, 264558	
627	80255457 (1253, 1254)	Novel Protein sim. GBank gij3098418 (AF040944) - P140 [Mus musculus]		UNCLASSIFIED	18108394, 264112, 264593, 265022, 264635	
628	80077096 (1255, 1256)	Novel Protein sim. GBank gij1711543[sp]P50526[SSP1_SCHPO - SERINE/THREONINE-PROTEIN KINASE SSP1]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264600	
629	79851602 (1257, 1258)	Novel Protein sim. GBank gij1143204 (U34305) - ORF2; Method: conceptual translation supplied by author. [Shigella sonnei]		isomerase	264906, 264907	
630	39565156 (1259, 1260)	Novel Protein sim. GBank gij3236368 (AF064748) - S3-12 [Mus musculus]		UNCLASSIFIED	264490	
631	20598718 (1261, 1262)	Novel Protein sim. GBank gij140687[sp]P11666[YGGB_ECOLI - HYPOTHETICAL 30.9 KD PROTEIN IN SBM-FBA INTERGENIC REGION (ORF 4) (F286)]			263978	
632	27843880 (1263, 1264)			UNCLASSIFIED	264906, 264600, 264605, 264769, 264689, 264486	
633	80477772 (1265, 1266)			UNCLASSIFIED	264769	
634	17938806 (1267, 1268)				265019	
635	79574506 (1269, 1270)			UNCLASSIFIED	264689	
636	79910981 (1271, 1272)			UNCLASSIFIED	264596, 264762, 264693	

637	82455786 (1273, 1274)	Novel Protein sim. GBank gi 2326739 emb CAB10953  - (298268) recN [Mycobacterium tuberculosis]		nuclease	264906, 264907, 264510, 264511, 264601, 264602, 264603, 264604, 264605, 18108351, 264762, 264766, 264687, 264769, 264688, 35695917, 264693, 264634, 264638, 264639, 264559, 18108385, 264636
638	14997457 (1275, 1276)	Novel Protein sim. GBank gi 4678662 emb CAB41074.1  - (AL049645) putative large ATP-binding protein [Streptomyces coelicolor]			
639	80204210 (1277, 1278)	Novel Protein sim. GBank gi 4589628 dbj BAA76836.1  - (AB023209) KIAA0992 protein [Homo sapiens]	struct		264112, 263974
640	17929579 (1279, 1280)	Novel Protein sim. GBank gi 1432083 (U60981) - homolog to Skp1p, an evolutionarily conserved kinetochore protein in budding yeast [Arabidopsis thaliana]	Contains protein domain (PF01466) - Skp1 family	(n)polymerase	265009, 265010
641	79636398 (1281, 1282)			UNCLASSIFIED	264693
642	18998737 (1283, 1284)			UNCLASSIFIED	264565
643	81516220 (1285, 1286)			UNCLASSIFIED	264906, 264908, 264758, 264288, 264632, 264635, 264639, 264564
644	11751367 (1287, 1288)			UNCLASSIFIED	264684
645	95010907 (1289, 1290)			UNCLASSIFIED	264906, 264762, 264693, 264639, 264559
646	80069083 (1291, 1292)				264595, 264566
647	80257085 (1293, 1294)	Novel Protein sim. GBank gi 4507613 ref NP_003738.1 pTNKS - TANKYRASE	Contains protein domain (PF00023) - Ank repeat	transcriptfactor	264909, 264591
648	80077428 (1295, 1296)	Novel Protein sim. GBank gi 1044963 bbs 169646 - protamine [Monodonta turbinata, gonads, Peptide, 106 aa]		UNCLASSIFIED	264600
649	80247447 (1297, 1298)			UNCLASSIFIED	263978
650	11798316 (1299, 1300)			UNCLASSIFIED	264686
651	11776932 (1301, 1302)	Novel Protein sim. GBank gi 1346916 sp P12283 PURA_ECOLI - ADENYLOSUCCINATE SYNTHETASE (IMP--ASPARTATE LIGASE)			264602, 264638
652	85516704 (1303, 1304)			UNCLASSIFIED	264905, 264907, 264909, 263978, 264637
653	82124947 (1305, 1306)	Novel Protein sim. GBank gi 1722977 sp Q10638 Y03C_MYCTU - HYPOTHETICAL 82.8 KD PROTEIN CY130.12C		UNCLASSIFIED	22278996, 264510, 264511, 264512, 264593, 21906754, 264603, 264760, 18108376, 264556
654	95010589 (1307, 1308)			UNCLASSIFIED	264906, 264595, 264632
655	78320692 (1309, 1310)	Novel Protein sim. GBank gi 130327 sp P26647 PLSC_ECOLI - 1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (1-AGP ACYLTRANSFERASE) (1-AGPAT) (LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE) (LPAAT)	Contains protein domain (PF01553) - Acyltransferase	transferase	264592
656	80416739 (1311, 1312)				
657	20611010 (1313, 1314)			UNCLASSIFIED	264602, 264605, 264766, 264691
				UNCLASSIFIED	264557, 264558

658	87761815 (1315, 1316)	Novel Protein sim. GBank gjl5689493[dbj][BAA483030.1] - (AB029001) KIAA1078 protein [Homo sapiens]		UNCLASSIFIED	22278996, 60432049, 29331822, 29331824, 29331828, 265007, 265009, 33657402, 33657084, 265017, 264448, 21906765, 21906766, 263967, 20281149, 18108370, 18108374, 264482
659	87718663 (1317, 1318)	Novel Protein sim. GBank gjl2137872[pir][48724 - zinc finger protein PFZ - mouse]	Contains protein domain (PF00086) - Zinc finger, C2H2 type	transcriptfactor	22278999, 60432049, 66714117, 29331827, 265007, 264766, 56181562, 18108359, 18108365, 18108370, 18108381
660	81897922 (1319, 1320)	Novel Protein sim. GBank gjl134180[spP15401][SACY_BACSU - LEVANSUCRASE AND SUCRASE SYNTHESIS OPERON ANTI-TERMINATOR]	Contains protein domain (PF00874) - Transcriptional antiterminator bglG family	UNCLASSIFIED	264757
661	80028023 (1321, 1322)	Novel Protein sim. GBank gjl4545229[gb][AAD22450.1][AF116183] SecA homolog [Actinobacillus actinomycetemcomitans]		UNCLASSIFIED	264603, 264604, 264605, 32833986, 18108376, 264636, 18108387, 22279000
662	20463731 (1323, 1324)	Novel Protein sim. GBank gjl16841[spP21640][COBJ_PSEDE - PRECORRIN-3B C17 METHYLTRANSFERASE (PRECORRIN-3 METHYLTRANSFERASE) (PRECORRIN-3 METHYLTRANSFERASE)]		UNCLASSIFIED	264605
663	20628080 (1325, 1326)	Novel Protein sim. GBank gjl5689250[dbj][BAA82881.1] - (AB024335) similar to orf5 [Comamonas testosteroni]		dehydrogenase	264605
664	80508512 (1327, 1328)	Novel Protein sim. GBank gjl1652848[dbj][BAA17766] - (D90909) DNA photolyase [Synecocystis sp.]		UNCLASSIFIED	264769
665	80079053 (1329, 1330)	Novel Protein sim. GBank gjl16841[spP21640][COBJ_PSEDE - PRECORRIN-3B C17 METHYLTRANSFERASE (PRECORRIN-3 METHYLTRANSFERASE) (PRECORRIN-3 METHYLTRANSFERASE)]		isomerase	264600
666	78603142 (1331, 1332)	Novel Protein sim. GBank gjl3261829[emb][CAB10927] - (Z98260) hypothetical protein Rv1230c [Mycobacterium tuberculosis]		glycoprotein	264907, 265007
667	94631802 (1333, 1334)	Novel Protein sim. GBank gjl5688851[dbj][BAA82702.1] - (AB017438) Orf5 [Streptomyces coelicolor]		UNCLASSIFIED	264689, 264602, 264593
668	82051891 (1335, 1336)	Novel Protein sim. GBank gjl3581853[emb][CAA20809] - (AL031541) 50S ribosomal protein L20 [Streptomyces coelicolor]	Contains protein domain (PF00453) - Ribosomal protein L20	ribosomalprot	264905, 264906, 264908, 264600, 264601, 264603, 264605, 264760, 264689, 264636, 264638, 264639
669	12967154 (1337, 1338)	Novel Protein sim. GBank gjl2582531 (AF026444) - 2-isopropylmalate synthase [Streptomyces coelicolor]		UNCLASSIFIED	264637
670	80238549 (1339, 1340)	Novel Protein sim. GBank gjl2582531 (AF026444) - 2-isopropylmalate synthase [Streptomyces coelicolor]		synthase	264905, 264906, 264908, 264601, 264762, 264766, 264689, 264638, 18108385, 264486
671	78601368 (1341, 1342)	Novel Protein sim. GBank gjl2114430 (U92703) - Oil-1/EBF-like-3 transcription factor [Mus musculus]	Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	264690, 264692, 264693, 264636, 18108387
672	79834371 (1343, 1344)	Novel Protein sim. GBank gjl2114430 (U92703) - Oil-1/EBF-like-3 transcription factor [Mus musculus]		transcriptfactor	264910, 265017
673	82285798 (1345, 1346)	Novel Protein sim. GBank gjl4589285[gb][AAD26430.1][AF135154] ferric alkaline siderophore receptor [Bordetella pertussis]			264759
674	78189259 (1347, 1348)			UNCLASSIFIED	264629

675	87895870 (1349, 1350)	Novel Protein sim. GBank gi 4980755 gb AAD35347.1 AE001708 D- alanine-D-alanine ligase [Thermotoga maritima]	Contains protein domain (PF01820) - D-ala D-ala ligase	UNCLASSIFIED	264488, 22278999, 66714117, 264508, 264511, 265008, 60433438, 264600, 264601, 264602, 264603, 264604, 264605, 264762, 264687, 264769, 60431602, 18108374, 264636, 264638 265010
676	78899607 (1351, 1352)	Novel Protein sim. GBank gi 1723566 sp Q10479 YDF7_SCHPO - PUTATIVE GLUCOSYLTRANSFERASE C17C9.07		ATPase-associated	264591, 264632
677	21644312 (1353, 1354)	Novel Protein sim. GBank gi 687208 (U03976) - dynein heavy chain isotype 5C [Tripleneustes gratilla]			
678	84225200 (1355, 1356)	Novel Protein sim. GBank gi 1586274 pr 2203365A - laminin alpha5 [Mus musculus]	Contains protein domain (PF00053) - Laminin EGF-like (Domains III and V)		264758, 264682, 264557
679	78868855 (1357, 1358)	Novel Protein sim. GBank gi 3928723 emb CAA22219  - (AL034355) putative ABC transporter [Streptomyces coelicolor]		UNCLASSIFIED	22278996, 264693
680	20726424 (1359, 1360)	Novel Protein sim. GBank gi 5174493 ref NP_006050.1 pLAMC - laminin, gamma 3	Contains protein domain (PF00053) - Laminin EGF-like (Domains III and V)		264600, 264602 264102, 264907, 264908, 265006, 264693, 263972, 83373044, 264566
682	11392476 (1363, 1364)	Novel Protein sim. GBank gi 4758208 ref NP_004081.1 pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	UNCLASSIFIED	264595 264634
683	80083680 (1365, 1366)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]			264605
685	80246735 (1369, 1370)				264909, 263967, 263981
686	78208608 (1371, 1372)				264631
687	80085629 (1373, 1374)				264693, 264635
688	78853412 (1375, 1376)	Novel Protein sim. GBank gi 2688962 (AF027768) - LspA [Serratia marcescens]		peptidase	264907, 264638
689	88064256 (1377, 1378)	Novel Protein sim. GBank gi 3046931 (AF049330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264906, 264907, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906766, 21906769, 265021, 18108361, 263974, 18108379, 264557, 18108385, 22279002 264510, 264511, 264764, 264769
690	80389750 (1379, 1380)	Novel Protein sim. GBank gi 2498941 sp Q15428 SP62_HUMAN - SPLICEOSOME ASSOCIATED PROTEIN 62 (SAP 62) (SF3A66)		UNCLASSIFIED	264757
691	81854392 (1381, 1382)			UNCLASSIFIED	55812038, 55811957, 265018, 55811150, 18108351, 264908, 60431528, 264594
692	83608936 (1383, 1384)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	264635
693	78586116 (1385, 1386)	Novel Protein sim. GBank gi 854065 emb CAA56337  - (X83413) U68 [Human herpesvirus 6]			22278996, 264510, 264602, 264603, 264762, 264687, 264769, 264693
694	82455983 (1387, 1388)	Novel Protein sim. GBank gi 267327 sp Q01033 VG48_HSVSA - HYPOTHETICAL GENE 48 PROTEIN			

695	94147849 (1389, 1390)	Novel Protein sim. GBank gij4468339[embjCAB38059.1] - (AJ010901) MUC4 [Homo sapiens]	Contains protein domain (PF00094) - von Willebrand factor type D domain	UNCLASSIFIED	58182575, 264509, 264905, 264907, 26331830, 264908, 264909, 264511, 265007, 264910, 264758, 264764, 264288, 65274791
696	79830882 (1391, 1392)	Novel Protein sim. GBank gij2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - ABC transporter	transport	264905, 264595
697	11767889 (1393, 1394)	Novel Protein sim. GBank gij1731343[spjQ10894]YY25_MYCTU - HYPOTHETICAL 24.4 KD PROTEIN CY49.25	Contains protein domain (PF01836) - Transposase	UNCLASSIFIED	264682
698	66695862 (1395, 1396)			UNCLASSIFIED	264688, 35695917
699	79582558 (1397, 1398)			UNCLASSIFIED	264682
700	79639098 (1399, 1400)			UNCLASSIFIED	264693
701	80230242 (1401, 1402)	Novel Protein sim. GBank gij1001236[dbjBAA10477] - (D54003) hypothetical protein [Synechocystis sp.]		UNCLASSIFIED	264488, 264510, 264511, 264602, 264605, 264689
702	79814789 (1403, 1404)			UNCLASSIFIED	264909
703	20446820 (1405, 1406)	Novel Protein sim. GBank gij2498935[spjQ46338]SOXG_CORSP - SARCOSINE OXIDASE GAMMA SUBUNIT	oxidase	UNCLASSIFIED	264604
704	94312224 (1407, 1408)	Novel Protein sim. GBank gij3150513 (AF067219) - contains similarity to the ketch/MIPP family [Caenorhabditis elegans]	Contains protein domain (PF01344) - Ketch motif	UNCLASSIFIED	264288, 56181562, 33657109, 264629, 55811576
705	17932141 (1409, 1410)	Novel Protein sim. GBank gij421091[prjIS30730] - hypothetical protein o206 - Escherichia coli		UNCLASSIFIED	265006
706	20288062 (1411, 1412)	Novel Protein sim. GBank gij3024872[spjQ35790]Y074_SYNY3 - HYPOTHETICAL 52.8 KD PROTEIN SLR0074			264600
707	20638065 (1413, 1414)	Novel Protein sim. GBank gij3420608[gbjAAC31907.1] - (AF075709) ABC transporter ATP-binding subunit [Pseudomonas putida]	transport		264603
708	20708282 (1415, 1416)	Novel Protein sim. GBank gij3649741[embjCAA03985] - (AJ000281) mucin [Homo sapiens]	struct		264601, 264692
709	88001439 (1417, 1418)	Novel Protein sim. GBank gij3080425[embjCAA18744.1] - (AL022604) putative protein [Arabidopsis thaliana]			18108398, 264637, 264908, 264909
710	11356683 (1419, 1420)				264369
711	17931418 (1421, 1422)	Novel Protein sim. GBank gij4758686[refjNP_002323.1]pLRP1 - low density lipoprotein related protein 1 (alpha-2-macroglobulin receptor)	Contains protein domain (PF00058) - apolipoprotein repeat class B		265019
712	80258164 (1423, 1424)				264591
713	79263126 (1425, 1426)	Novel Protein sim. GBank gij1703268[spjQ11056]AMIZ_MYCTU - PUTATIVE AMIDASE CY50.19C	hydrolase		264905, 264907
714	27647651 (1427, 1428)	Novel Protein sim. GBank gij4502351[refjNP_001692.1]pBAAT - bile acid Coenzyme A: amino acid N-acyltransferase, glycine N-choyltransferase			264508, 264555

715	79639423 (1429, 1430)	Novel Protein sim. GBank gj1780035 (AE000352) - ori, hypothetical protein [Escherichia coli]		UNCLASSIFIED	264907	
716	79559072 (1431, 1432)				264692	
717	79491842 (1433, 1434)	Novel Protein sim. GBank gj2494074 [sp]P55653 [GABD_RHISN - PROBABLE SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+)] (SSDH)		dehydrogenase	264636	
718	94318658 (1435, 1436)	Novel Protein sim. GBank gj3873679 [emb]CAA94886] - (Z71178) similar to pro-collagen domains; cDNA EST EMBL:D27978 comes from this gene; cDNA EST EMBL:D27977 comes from this gene; cDNA EST EMBL:D34199 comes from this gene; cDNA EST EMBL:D64392 comes from this gene; cDNA EST EMBL... (Z95387) hypothetical protein Rv2811c [Mycobacterium tuberculosis]	Contains protein domain (PF000093) - von Willebrand factor type C domain	kinase	18108392, 22278994, 22278998, 265008, 265018, 264681, 18108354, 264684, 264685, 264686, 264687, 264689, 21906769, 18108361, 264691, 264692, 55810764, 264635, 18108381, 18108382, 83373044, 18108388	
719	17679564 (1437, 1438)	Novel Protein sim. GBank gj2104302 [emb]CAB08631] - (Z95387) hypothetical protein Rv2811c [Mycobacterium tuberculosis]		UNCLASSIFIED	265011	
720	79841684 (1439, 1440)				264908	
721	15020180 (1441, 1442)	Novel Protein sim. GBank gj1235303 [sp]P04929 [HRPX_PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR		UNCLASSIFIED	264629	
722	9862603 (1443, 1444)	Novel Protein sim. GBank gj498253 (U02372) - integrase [Vibrio cholerae]			264910	
723	18755599 (1445, 1446)	Novel Protein sim. GBank gj2253054 [emb]CAB10705] - (Z97559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264691	
724	10126494 (1447, 1448)	Novel Protein sim. GBank gj4083015 (AF083061) - protease PrtA [Pseudomonas fluorescens]	Contains protein domain (PF00353) - protease Hemolysin-type calcium-binding proteins		264909	
725	78878679 (1449, 1450)			UNCLASSIFIED	264905, 264907	
726	13086282 (1451, 1452)			UNCLASSIFIED	264636	
727	13522872 (1453, 1454)				264634	
728	20268471 (1455, 1456)	Novel Protein sim. GBank gj2633910 [emb]CAB13411] - (Z99112) similar to hypothetical proteins [Bacillus subtilis]			264567	
729	11293753 (1457, 1458)			UNCLASSIFIED	264490	
730	18900373 (1459, 1460)	Novel Protein sim. GBank gj2494680 [sp]Q45291 [GALE_BRELA - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)		isomerase	264564	
731	80058750 (1461, 1462)	Novel Protein sim. GBank gj1146192 (L47838) - putative [Bacillus subtilis]		UNCLASSIFIED	264605	
732	80258175 (1463, 1464)	Novel Protein sim. GBank gj1168396 [sp]P46681 [AIP2_YEAST - ACTIN INTERACTING PROTEIN 2		struct	264591, 264594, 264595	
733	20446839 (1465, 1466)			UNCLASSIFIED	264604	
734	20435987 (1467, 1468)	Novel Protein sim. GBank gj3184080 [emb]CAA19336] - (AL023781) hypothetical protein [Schizosaccharomyces pombe]		ubiquitin	264604	

735	11607959 (1469, 1470)	Novel Protein sim. GBank gi401582 sp P27432 YICE_ECOLI - HYPOTHETICAL 48.9 KD PROTEIN IN GLTS-SELG INTERGENIC REGION			264594
736	10878734 (1471, 1472)	Novel Protein sim. GBank gi400831 sp P31135 POTH_ECOLI - PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTH	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	264636
737	78945340 (1473, 1474)		Contains protein domain (PF00815) - Regulator of G protein signaling domain	UNCLASSIFIED	265020
738	17895353 (1475, 1476)	Novel Protein sim. GBank			265008
739	79833670 (1477, 1478)	gi2506867 sp P33225 TORA_ECOLI - TRIMETHYLAMINE- N-OXIDE REDUCTASE PRECURSOR (TMAO REDUCTASE) (TRIMETHYLAMINE OXIDASE)		oxidase	264910
740	19881557 (1479, 1480)	Novel Protein sim. GBank gi3261828 emb CAB10925  - (Z98260) mtp [Mycobacterium tuberculosis]			264907, 264764, 264634, 264637
741	78827273 (1481, 1482)	Novel Protein sim. GBank gi3877494 emb CAA88472.1  - (Z48583) ATP binding protein with similarity to the CDC48/PAS1/SEC18 family; cDNA EST EMBL:D65037 comes from this gene; cDNA EST EMBL:D68340 comes from this gene; cDNA EST EMBL:D65048 comes from this gene; cDNA EST EMBL:D6845...	Contains protein domain (PF01883) - Domain of unknown function	UNCLASSIFIED	264689, 35696286, 264510, 264908, 18108362
742	82393795 (1483, 1484)	Novel Protein sim. GBank gi127420 sp P19888 MTBA_BACAR - MODIFICATION METHYLASE BANI (CYTOSINE-SPECIFIC METHYLTRANSFERASE BANI) (M.BANI)		UNCLASSIFIED	29331822, 264910, 264762
743	82300051 (1485, 1486)	Novel Protein sim. GBank gi127420 sp P19888 MTBA_BACAR - MODIFICATION METHYLASE BANI (CYTOSINE-SPECIFIC METHYLTRANSFERASE BANI) (M.BANI)	Contains protein domain (PF00145) - C-5 cytosine-specific DNA methylase	UNCLASSIFIED	264488, 264259, 264508, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 265008, 265009, 264910, 264591, 264596, 264759, 265010, 265011, 18108351, 264763, 264288, 264766, 264768, 264693, 18108370, 264629, 18108372, 264630, 264631, 264834, 264558, 18108385, 264482, 264584, 264567
744	80230421 (1487, 1488)				18108397, 264511, 264690, 264628, 264638, 264692, 264639, 264766
745	9841963 (1489, 1490)	Novel Protein sim. GBank gi178921 pir IS04846 - UDP-N- acetylmutamoylalanine-D-glutamate-2, 6-diaminopimelate--O- alanine-D-alanine ligase (EC 6.3.2.15) precursor - Escherichia coli		glycoprotein	264906
746	11073229 (1491, 1492)	Novel Protein sim. GBank gi3386354 (AF074705) - pyococcal synthetase [Pseudomonas aeruginosa]		synthase	264600
747	94322044 (1493, 1494)	Novel Protein sim. GBank gi288741 dbj BA24848  - (AB007878) KIAA0418 [Homo sapiens]	Contains protein domain (PF00018) - SH3 domain	oxidase	66714117, 264905, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 265011, 264681, 264288, 264766, 264687, 264768, 264769, 21906768, 35695917, 264691, 264693, 264628, 264634, 264635, 264639, 56182323, 83373044
748	11617923 (1495, 1496)				264690

749	20469119 (1497, 1498)	Novel Protein sim. GBank gi 1169727 sp P44948 FPG_HAEIN - FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (FAPY- DNA GLYCOSYLASE)		UNCLASSIFIED	264604
750	20298427 (1499, 1500)	Novel Protein sim. GBank gi 5360068 gb AAD42851.1 AF15968 - (AF159689) serine/threonine kinase PKN3 [Myxococcus xanthus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain		264600
751	21636169 (1501, 1502)	Novel Protein sim. GBank gi 1168662 sp P44426 BIOA_HAEIN - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7,8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)	UNCLASSIFIED	264605, 264559	
752	82450386 (1503, 1504)	Novel Protein sim. GBank gi 2651530 sp P32399 YHGE_BACSU - HYPOTHETICAL 84.1 KD PROTEIN IN HEMY-GLTT INTERGENIC REGION (ORF8)	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal phosphate	264508, 264907, 264510, 265011, 264762, 264689, 35695855, 264638, 18108387	
753	80508718 (1505, 1506)	Novel Protein sim. GBank gi 2651530 sp P32399 YHGE_BACSU - HYPOTHETICAL 84.1 KD PROTEIN IN HEMY-GLTT INTERGENIC REGION (ORF8)	UNCLASSIFIED	264909, 264600, 264602, 264604, 264760, 264769, 264634	
754	95083741 (1507, 1508)	Novel Protein sim. GBank gi 3449276 emb CAA20420  - (AL031317) putative dehydrogenase [Streptomyces coelicolor]	UNCLASSIFIED	264508, 264906, 264907, 264908, 264909, 264759, 264602, 264764, 264769, 264628, 264629, 264630, 264632, 264634, 264635, 264637, 264638, 83373044, 18108385	
755	80185449 (1509, 1510)	Novel Protein sim. GBank gi 3449276 emb CAA20420  - (AL031317) putative dehydrogenase [Streptomyces coelicolor]	UNCLASSIFIED	264448, 264690	
756	94631686 (1511, 1512)	Novel Protein sim. GBank gi 4360331 emb CAB40107.1  - (AJ001206) putative glycogen debranching enzyme [Streptomyces coelicolor]	UNCLASSIFIED	264682, 264685	
757	79468533 (1513, 1514)	Novel Protein sim. GBank gi 2911858 (AF047659) - No definition line found [Caenorhabditis elegans]	amylase	265007, 18108387, 265007, 18108387	
758	78963176 (1515, 1516)	Novel Protein sim. GBank gi 3451312 emb CAA20449  - (AL031324) membrane atpase [Schizosaccharomyces pombe]	UNCLASSIFIED	264684, 264686	
759	79475667 (1517, 1518)	Novel Protein sim. GBank gi 3327158 dbj BAA31647  - (AB014572) KIAA0672 protein [Homo sapiens]	Contains protein domain (PF00122) - E1-E2 ATPase	29331822, 264908, 52844045, 56182435, 60170831, 21906754, 265017, 265019, 264681, 264687, 264688, 21906768, 21908768, 265020, 265021, 265022, 264635, 22279000	
760	87628888 (1519, 1520)	Novel Protein sim. GBank gi 4981266 gb AAD35822.1 AE00174 - (AE001744) lipopolysaccharide core biosynthesis protein KdtB [Thermotoga maritima]	transport	264766	
761	79877966 (1521, 1522)	Novel Protein sim. GBank gi 3327158 dbj BAA31647  - (AB014572) KIAA0672 protein [Homo sapiens]	UNCLASSIFIED	264907, 264593, 265020	
762	80023563 (1523, 1524)	Novel Protein sim. GBank gi 4981266 gb AAD35822.1 AE00174 - (AE001744) lipopolysaccharide core biosynthesis protein KdtB [Thermotoga maritima]	UNCLASSIFIED	264600	
763	20294813 (1525, 1526)	Novel Protein sim. GBank gi 4981266 gb AAD35822.1 AE00174 - (AE001744) lipopolysaccharide core biosynthesis protein KdtB [Thermotoga maritima]			
764	39515024 (1527, 1528)	Novel Protein sim. GBank gi 4981266 gb AAD35822.1 AE00174 - (AE001744) lipopolysaccharide core biosynthesis protein KdtB [Thermotoga maritima]		264603	



765	80025347 (1529, 1530)	Novel Protein sim. GBank gij3845093 (AE001371) - erythrocyte membrane protein PIEMP3 [Plasmodium falciparum]		struct	264905, 264906, 264594, 264686, 33657023
766	82417404 (1531, 1532)	Novel Protein sim. GBank gij541121 [pir] [S40827 - hypothetical protein o300 - Escherichia coli]		UNCLASSIFIED	264605, 264762, 18108374
767	10296742 (1533, 1534)	Novel Protein sim. GBank gij2982501 [emb] [CAA06184] - (AJ004832) neuropathy target esterase [Homo sapiens]		UNCLASSIFIED	264592, 264595
768	79416080 (1535, 1536)	Novel Protein sim. GBank gij283437 [pir] [S27850 - hypothetical protein - Trypanosoma cruzi (fragment)]		esterase	55810764, 264559
769	80086554 (1537, 1538)	Novel Protein sim. GBank gij4769004 [gb] [AAD29715.1] [AF14059 - (AF140598) ring-box protein 1 [Homo sapiens]]		UNCLASSIFIED	264905, 264907, 264828, 264909, 265010, 264766, 264628, 264629, 264634, 264636, 264555
770	80417847 (1539, 1540)	Novel Protein sim. GBank gij5531324 [emb] [CAB51045.1] - (AJ009579) putative alkane 1-monooxygenase [Pseudomonas fluorescens]	Contains protein domain (PF000097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	56182575, 35696286, 264259, 66714117, 264107, 66712502, 56182435, 264112, 55812038, 87168559, 264288, 21908766, 33657023, 65274620, 65274791, 18108381
771	95329509 (1541, 1542)	Novel Protein sim. GBank gij4505481 [ref] [NP_003624.1] [pNRPB - nuclear restricted protein, BTB domain-like (brain)]		UNCLASSIFIED	264910
772	78971362 (1543, 1544)	Novel Protein sim. GBank gij1144520 (U34956) - phosphoribosylformylglycinamide synthase [Mycobacterium tuberculosis]		UNCLASSIFIED	265020
773	78945363 (1545, 1546)	Novel Protein sim. GBank gij4155447 (AE001517) - proline/betaine transporter [Helicobacter pylori J99]		UNCLASSIFIED	264909
774	79856129 (1547, 1548)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]	Contains protein domain (PF01006) - Hepatitis C virus non-structural protein NS4a	UNCLASSIFIED	264555
775	20620141 (1549, 1550)	Novel Protein sim. GBank gij136748 [sp] [P10905] [UGPA, ECOLI - SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPA]	Contains protein domain (PF01344) - Kelch motif	UNCLASSIFIED	265019
776	78942693 (1551, 1552)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		protease	21906754, 265020, 80170615, 264691
777	79960378 (1553, 1554)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		UNCLASSIFIED	264511
778	20691310 (1555, 1556)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		UNCLASSIFIED	264603
779	80054024 (1557, 1558)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		synthase	264907, 264600, 264601, 264602, 264603, 264604, 264605, 264486
780	95288987 (1559, 1560)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]			264905, 264907, 265010, 264600, 264601, 18108362, 18108374, 264556
781	80250049 (1561, 1562)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		UNCLASSIFIED	264605
782	8758529 (1563, 1564)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		UNCLASSIFIED	265020
783	16410791 (1565, 1566)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		UNCLASSIFIED	264635, 33657023, 29331828, 265017, 264565, 264566
784	80051197 (1567, 1568)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	struct	35696052, 264604
785	56073541 (1569, 1570)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		transport	264603
786	20438842 (1571, 1572)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]			
787	80258364 (1573, 1574)	Novel Protein sim. GBank gij3451335 (AC005525) - F22162.1 [Homo sapiens]		UNCLASSIFIED	264593

788	80507844 (1575, 1576)	Novel Protein sim. GBank gij2746079 (AF015310) - BTH1 [Brassica napus]		synthase	264909, 264602, 264603, 264769, 264638
789	17294715 (1577, 1578)	Novel Protein sim. GBank gij2351849 (U83357) - 40 kDa heat shock chaperone protein [Halobacterium salinarum]		UNCLASSIFIED	265007
790	86284408 (1579, 1580)	Novel Protein sim. GBank gij5706378 [dbj]BAA83099.1] - (AB026118) MAL T1 [Homo sapiens]	Contains protein domain (PF000047) - Immunoglobulin domain	glycoprotein	22278999, 29331824, 264828, 87168559, 265018, 21906765, 21906767, 21906768, 21906769, 265020, 264692, 22279000, 264563
791	94651627 (1581, 1582)	Novel Protein sim. GBank gij5689948 [emb]CAB51985.1] - (AL109663) putative isoleucyl-tRNA synthetase [Streptomyces coelicolor A3(2)]			264601, 264605, 264636
792	80058786 (1583, 1584)	Novel Protein sim. GBank gij393184 (L02375) - S-antigen [Plasmodium falciparum]		struct	265021, 264631, 264635, 264558
793	79638730 (1585, 1586)	Novel Protein sim. GBank gij1345408 [dbj]BAA05046] - (D26046) AT motif-binding factor [Mus musculus]	Contains protein domain (PF000046) - Homeobox domain	homeobox	264693
794	81839294 (1587, 1588)	Novel Protein sim. GBank gij105884 [pir]S24023 - dopamine receptor D4 - human (fragment)		UNCLASSIFIED	264603, 264604, 264910, 264762, 264906, 264639, 264909, 264757
795	80074988 (1589, 1590)	Novel Protein sim. GBank gij1877334 [emb]CAB07082] - (Z92771) birA [Mycobacterium tuberculosis]		carboxylase	264488, 35696052, 264905, 264907, 265010, 35696423, 264636
796	86669451 (1591, 1592)				60432229, 55811150, 264630, 264637, 264565
797	87771781 (1593, 1594)	Novel Protein sim. GBank gij2995447 [emb]CAA71519] - (Y10495) CDV-1R protein [Mus musculus]		struct	22278998, 264093, 264094, 66714117, 21906767, 21906769, 265020, 265022
798	79865209 (1595, 1596)	Novel Protein sim. GBank gij4467250 [emb]CAB37575] - (AL035569) probable Glu-tRNA Gln amidotransferase subunit [Streptomyces coelicolor]		transcript factor	264687, 264768, 264693
799	79557816 (1597, 1598)			hydrolase	264908, 264910, 264636, 264638
800	79970189 (1599, 1600)			UNCLASSIFIED	264488
801	80499399 (1601, 1602)	Novel Protein sim. GBank gij2791517 [emb]CAA18054] - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]		transport	264508, 264511, 265006, 265009, 264769, 264567, 264486
802	79834588 (1603, 1604)	Novel Protein sim. GBank gij4887211 [gb]AAD32237.1] [AF14744] - (AF147449) penicillin binding protein 1B [Pseudomonas aeruginosa]			264905, 264693
803	20467520 (1605, 1606)			struct	
804	10174239 (1607, 1608)	Novel Protein sim. GBank gij1176152 [sp]P44507 [YHAD_HAEIN - HYPOTHETICAL PROTEIN H10091]		Kinase	264510
805	79599993 (1609, 1610)				
806	80484113 (1611, 1612)	Novel Protein sim. GBank gij2784612 [emb]CAA04683] - (AJ001330) ornithine transcarbamoylase [Lactobacillus sakei]	Contains protein domain (PF00185) - Aspartate/ornithine carbamoyltransferase	transferase	264508, 264769
807	80381812 (1613, 1614)	Novel Protein sim. GBank gij2833311 [sp]Q21828 [YNF0_CAEEL - HYPOTHETICAL 18.9 KD PROTEIN R07E5.13 IN CHROMOSOME III]			264764

808	35106617 (1615, 1616)	Novel Protein sim. GBank gi 3913092 sp Q46170 ARCD_CLOPE - ARGININE/ORNITHINE ANTIporter		transport	264909, 264602, 21905764, 18108374
809	81454254 (1617, 1618)	Novel Protein sim. GBank gi 3913016 sp P74309 ALF1_SYNY3 - FRUCTOSE- BISPHOSPHATE ALDOLASE CLASS I (FBP ALDOLASE)	Contains protein domain (PF00274) - Fructose-bisphosphate aldolase class-I	UNCLASSIFIED	264508, 264906, 264909, 265007, 264910, 264758, 264600, 264602, 264603, 264605, 264687, 264769, 264689, 264636, 264486
810	80192761 (1619, 1620)	Novel Protein sim. GBank gi 401472 sp P30863 YAFB_ECOLI - HYPOTHETICAL OXIDOREDUCTASE IN ASPU-MLTD INTERGENIC REGION	Contains protein domain (PF00248) - Aldol/keto reductase family	reductase	284369
811	80079280 (1621, 1622)			UNCLASSIFIED	264558
812	10297654 (1623, 1624)			UNCLASSIFIED	264692
813	79612280 (1625, 1626)			UNCLASSIFIED	264906
814	80473427 (1627, 1628)	Novel Protein sim. GBank gi 146168 (J01617) - glutamyl- tRNA synthetase [Escherichia coli]		synthase	264905, 264602, 264605, 264682, 264687, 264769, 264636
815	95419513 (1629, 1630)	Novel Protein sim. GBank gi 4589652 dbj BAA76848.1 - (A8023221) KIAA1004 protein [Homo sapiens]		UNCLASSIFIED	264486, 22278998, 22278999, 29331822, 29331824, 29331825, 29331827, 29331828, 29146499, 264905, 264908, 265007, 33657402, 60433356, 60433438, 264758, 265011, 265017, 265018, 265019, 264369, 264288, 264885, 21906765, 21906767, 265020, 265021, 264692, 65274820, 33657109, 264628, 18108376, 264635, 264638, 60170394, 56182323, 264564
816	19881910 (1631, 1632)				264600
817	95293316 (1633, 1634)	Novel Protein sim. GBank gi 1781144 emb CAB06254 - (Z83866) hypothetical protein Rv3069 [Mycobacterium tuberculosis]		UNCLASSIFIED	264595
818	90938190 (1635, 1636)	Novel Protein sim. GBank gi 1477468 (U35244) - vacuolar protein sorting homolog v-ps33a [Rattus norvegicus]			65274572, 22278999, 60424269, 35696052, 55812038, 21906768, 55811957, 35695917, 33657023, 18108370, 18108374, 55810764, 35696423, 55811576, 264636
819	80254977 (1637, 1638)	Novel Protein sim. GBank gi 1001352 dbj BAA10839 - (D64006) ABC transporter [Synecocystis sp.]		transport	264565
820	80059688 (1639, 1640)	Novel Protein sim. GBank gi 586814 sp P37484 YYBT_BACSU - HYPOTHETICAL 74.3 KD PROTEIN IN RPLI-COTF INTERGENIC REGION		UNCLASSIFIED	264600, 264602, 264604
821	79762590 (1641, 1642)				
822	80215310 (1643, 1644)			UNCLASSIFIED	264910
823	94892299 (1645, 1646)	Novel Protein sim. GBank gi 3878400 emb CAA95828 - (Z71264) predicted using GeneFinder: Weak similarity to Mouse T-complex-associated-testes-expressed-1 protein (PIR Acc. No. A45841); cDNA EST EMBL:D32742 comes from this gene; cDNA EST EMBL:D33617 comes from this gene; cDNA EST...		struct	264510, 264594, 264637 264509, 264687, 264691
824	80411171 (1647, 1648)	Novel Protein sim. GBank gi 1370076 emb CAA66887 - (X98235) type I [Drosophila melanogaster]	Contains protein domain (PF01429) - Methyl-CpG binding domain		264910, 264763, 264769, 264693

825	20638600 (1649, 1650)	Novel Protein sim. GBank gi 3025132 sp P77391 YEAG_ECOLI - HYPOTHETICAL 74.5 KD PROTEIN IN GAPA-RND INTERGENIC REGION		UNCLASSIFIED	264592
826	11075047 (1651, 1652)	Novel Protein sim. GBank gi 3242281 emb CAA16669  - (AL021646) hypothetical protein Rv3202c [Mycobacterium tuberculosis]			264605
827	80054207 (1653, 1654)	Novel Protein sim. GBank gi 3417424 emb CAA20312  - (AL031261) putative transport protein [Schizosaccharomyces pombe]			264603
828	95106322 (1655, 1656)	Novel Protein sim. GBank gi 4336892 gb AAD17897  - (AF101361) Abnormal X segregation [Drosophila melanogaster]		UNCLASSIFIED	52645080, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264512, 264910, 264591, 264758, 264600, 264766, 264768, 21908768, 35695917, 264691, 264628, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264558, 264639, 83373044, 18108385, 264563, 264566, 264486
829	81742215 (1657, 1658)			UNCLASSIFIED	264758, 264634
830	20396091 (1659, 1660)	Novel Protein sim. GBank gi 3820539 (AF080002) - UDP-N- acetylmutamyl tripeptide synthetase MurC [Helicobacter mobilis]		UNCLASSIFIED	264603
831	87112435 (1681, 1682)			UNCLASSIFIED	66714117, 264910, 264639
832	18536322 (1663, 1664)	Novel Protein sim. GBank gi 1870004 emb CAB06855  - (Z92539) hypothetical protein Rv1024 [Mycobacterium tuberculosis]		UNCLASSIFIED	264906
833	20726654 (1665, 1666)	Novel Protein sim. GBank gi 2500056 sp Q46267 PFLA_CLOPA - PYRUVATE FORMATE-LYASE ACTIVATING ENZYME		UNCLASSIFIED	264602
834	21428762 (1667, 1668)	Novel Protein sim. GBank gi 2497531 sp Q46078 KPYK_CORGL - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	264600, 264602, 264769, 264689, 264636
835	94140482 (1669, 1670)				264768, 263994, 21906767, 264910, 264632, 264635, 264259, 264639, 264693, 83373044, 264758, 35696052, 22279002, 264508, 264905, 264906, 264448, 263972, 264908, 264909
836	66126552 (1671, 1672)	Novel Protein sim. GBank gi 699315 (U15184) - phosphate transport protein PSTC [Mycobacterium leprae]		transport	35695917, 264557
837	79450450 (1673, 1674)			UNCLASSIFIED	264595
838	79184203 (1675, 1676)	Novel Protein sim. GBank gi 728867 sp P40602 APG_ARATH - ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR		UNCLASSIFIED	264687
839	79641125 (1677, 1678)	Novel Protein sim. GBank gi 2496533 sp Q50598 MYCTU - HYPOTHETICAL 69.9 KD PROTEIN CY1A11.08		UNCLASSIFIED	264906
840	80058851 (1679, 1680)	Novel Protein sim. GBank gi 4557753 ref NP_000372.1 pMID1 - midline 1 protein (finger)	Contains protein domain (PF000097) - Zinc finger, C3HC4 type (RING finger)	interleukinrecept	264762, 264556

841	80376318 (1681, 1682)	Novel Protein sim. GBank gij139805[sp]P08045[XFIN_XENLA - XFIN PROTEIN	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264764
842	80078724 (1683, 1684)	Novel Protein sim. GBank gij2114321[dbj]BAA200371 - (D88733) membrane glycoprotein [Equine herpesvirus 1]	Contains protein domain (PF00569) - Zinc finger present in dystrophin, CBP/p300	UNCLASSIFIED	264905, 264908, 265008, 265009, 18108374, 56182323, 264558
843	87002847 (1685, 1686)	Novel Protein sim. GBank gij3882325[dbj]BAA34522.1  - (AB018345) KIAA0802 protein [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	struct	264091, 29331825, 264906, 264768, 264563
844	17941439 (1687, 1688)	Novel Protein sim. GBank gij2224721[dbj]BAA20844  - (AB002388) KIAA0390 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	265011
845	18346844 (1689, 1690)	Novel Protein sim. GBank gij625679[pri]A36929 - virulence regulatory protein VsrB - Pseudomonas solanacearum		kinase	264629
846	79863441 (1691, 1692)				284907
847	78695348 (1693, 1694)				264909
848	78489385 (1695, 1696)			UNCLASSIFIED	265020
849	79756367 (1697, 1698)			UNCLASSIFIED	264566
850	78817849 (1699, 1700)	Novel Protein sim. GBank gij3183245[sp]P78061[JC_ECOLI - PUTATIVE GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)	Contains protein domain (PF00120) - Glutamine synthetase	UNCLASSIFIED	264909
851	95320333 (1701, 1702)	Novel Protein sim. GBank gij5454130[ref]NP_006280.1[pTIN] - talin	Contains protein domain (PF01608) - VILWEQ domain		264488, 52644507, 264489, 18108398, 65274572, 56182575, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 20281171, 264490, 264259, 52645080, 29331822, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 29146498, 29146499, 264107, 264905, 264906, 264907, 264908, 52644045, 56182435, 265006, 265007, 265008, 265009, 264910, 60432229, 60431735, 60433356, 33657402, 60433438, 264595, 264758, 264759, 21906754, 33109954, 52644296, 265010, 265011, 87168559, 285017, 265018, 265019, 264760, 264761, 264762, 264681, 18108351, 264763, 264448, 264682, 264764, 264683, 18108354, 264288, 264369, 264685, 264766, 264687, 264768, 264769, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 29148629, 55811857, 35695917, 265020, 265021, 265022, 60170615, 52644150, 264691, 264692, 33657023, 264693, 263966, 33657109, 27486261, 27486262, 27486264, 27486265, 35695763, 60431602, 18108370, 20281069, 264629, 18108374, 18108376, 55811576, 35696423, 35695855, 264634, 264635, 264636, 264555, 60431850, 264556, 264691
852	10147366 (1703, 1704)				

853	13032587 (1705, 1706)	Novel Protein sim. GBank gij3402836jembj[CAA76082] - (Y16136) 2-enoate reductase [Moorella thermoacelica]		reductase	264636 264566
854	80052438 (1707, 1708)				
855	79641130 (1709, 1710)			UNCLASSIFIED	264692 264591
856	11594238 (1711, 1712)			UNCLASSIFIED	264630, 264634
857	79210165 (1713, 1714)				265008, 265009, 264601, 264602, 264603, 18108351
858	80248910 (1715, 1716)				264559
859	20296634 (1717, 1718)			UNCLASSIFIED	264489
860	80041749 (1719, 1720)			UNCLASSIFIED	33657023, 264630
861	65857045 (1721, 1722)				264600
862	80079467 (1723, 1724)				264488, 18108398, 35696286, 264259, 18108351, 264288, 265021
863	80579931 (1725, 1726)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 79, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	
864	94839904 (1727, 1728)	Novel Protein sim. GBank gij5689884jembj[CAB52047.1] - (AL108732) hypothetical protein [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01479) - S4 domain	UNCLASSIFIED	264259, 264112, 263974
865	80045310 (1729, 1730)	Novel Protein sim. GBank gij4557876jrefj[NP_000341.1]pABCR - ATP binding cassette transporter		transport	264288, 264557, 264558
866	80162031 (1731, 1732)				
867	80062402 (1733, 1734)				264605
868	10075384 (1735, 1736)			UNCLASSIFIED	264909
869	80062406 (1737, 1738)				264605, 264687, 18108374
870	80249651 (1739, 1740)	Novel Protein sim. GBank gij628660jpirj[IS37755 - Adenyltransferase - Escherichia coli]		transferase	264601, 264636
871	20378295 (1741, 1742)	Novel Protein sim. GBank gij1708180jspj[Q10602]HEMK_MYCTU - HEMK PROTEIN HOMOLOG		UNCLASSIFIED	264603
872	95197114 (1743, 1744)	Novel Protein sim. GBank gij1545959jembj[CAA67763] - (X99384) paladin [Mus musculus]		UNCLASSIFIED	35696286, 22278998, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265006, 264511, 264512, 265007, 265008, 265009, 264910, 264591, 60433356, 264596, 52846317, 87168474, 265010, 264602, 264603, 265017, 265018, 264605, 18108351, 264764, 264766, 264768, 52644229, 264769, 21906765, 265021, 264534, 264691, 52645129, 264628, 264629, 35696423, 65274791, 264631, 264632, 264635, 264636, 264556, 264637, 264638, 264639, 60432113, 22279000, 22279002, 264564
873	20189728 (1745, 1746)	Novel Protein sim. GBank gij4156104 (AE001569) - putative Outer membrane protein [Helicobacter pylori J99]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	264595

874	80077692 (1747, 1748)	Novel Protein sim. GBank gij134319 sp P07819 SCRB_BACSU - SUCROSE-6- PHOSPHATE HYDROLASE (SUCRASE) (INVERTASE)		UNCLASSIFIED	264600
875	86608446 (1748, 1750)	Novel Protein sim. GBank gij481000 p rij S37594 - mucin - human (fragment)			264259, 264448, 264288, 264557, 87168518
876	86465157 (1751, 1752)	Novel Protein sim. GBank gij3128283 (AF010496) - iron(ii) diclrate transport ATP-binding protein [Rhodobacter capsulatus]	Contains protein domain (PF00005) - ABC transporter	transport	264907, 264601, 264602, 264605, 265020, 60431602
877	87802548 (1753, 1754)	Novel Protein sim. GBank gij731074 sp P40349 URB1 - SIDEROPHORE BIOSYNTHESIS REGULATORY PROTEIN URBS1	Contains protein domain (PF00320) - GATA zinc finger	transcriptfactor	22278998, 264909, 264369
878	80187289 (1755, 1756)	Novel Protein sim. GBank gij1351614 sp Q09853 YAED_SCHPO - HYPOTHETICAL 181.5 KD PROTEIN C23D3.13C IN CHROMOSOME I		ATPase_associated	264369, 264555
879	94328962 (1757, 1758)	Novel Protein sim. GBank gij3875304 emb CAA88434  - (Z74030) predicted using Genefinder; cDNA EST EMBL: C07608 comes from this gene; cDNA EST EMBL: C09023 comes from this gene; cDNA EST yk505e9.3 comes from this gene; cDNA EST yk489h9.3 comes from this gene; cDNA EST yk489h9.5 com...			56182575, 29331824, 264508, 264906, 265018, 18108351, 264448, 264683, 21906766, 21906768, 60170615, 33657023, 65274620, 33657109, 18108374, 35695855, 264563
880	8491135 (1759, 1760)	Novel Protein sim. GBank gij137120 sp P11214 UROT_MOUSE - TISSUE PLASMINOGEN ACTIVATOR PRECURSOR (TPA) (T- PLASMINOGEN ACTIVATOR)	Contains protein domain (PF00051) - Kingle domain	cathepsin	264508
881	11290122 (1761, 1762)	Novel Protein sim. GBank gij2632098 emb CAA75667  - (Y15513) Prodes protein [Drosophila melanogaster]		UNCLASSIFIED	264508
882	11077011 (1763, 1764)	Novel Protein sim. GBank gij1155068 emb CAA64425  - (X94976) cell wall-plasma membrane linker protein [Brassica napus]		UNCLASSIFIED	264558
883	78582969 (1765, 1766)	Novel Protein sim. GBank gij1155068 emb CAA64425  - (X94976) cell wall-plasma membrane linker protein [Brassica napus]		UNCLASSIFIED	264688
884	13517921 (1767, 1768)	Novel Protein sim. GBank gij2078027 emb CAB08467  - (Z95208) hypothetical protein Rv2372c [Mycobacterium tuberculosis]		UNCLASSIFIED	264636
885	80052457 (1769, 1770)	Novel Protein sim. GBank gij2078027 emb CAB08467  - (Z95208) hypothetical protein Rv2372c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 18108362
886	11685136 (1771, 1772)	Novel Protein sim. GBank gij2685834 emb CAA15904  - (AL021006) sucA [Mycobacterium tuberculosis]			264690
887	94315307 (1773, 1774)	Novel Protein sim. GBank gij5689395 dbj BAA82981.1  - (AB028952) KIAA1029 protein [Homo sapiens]		dehydrogenase	35696052, 264906, 264600, 264603, 35695917, 35695855, 264636
888	10083399 (1775, 1776)	Novel Protein sim. GBank gij1861335 dbj BAA19365  - (AB001488) PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE [Bacillus subtilis]		UNCLASSIFIED	264908
889	20385917 (1777, 1778)	Novel Protein sim. GBank gij1861335 dbj BAA19365  - (AB001488) PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE [Bacillus subtilis]			264603
890	19904337 (1779, 1780)	Novel Protein sim. GBank gij854065 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]			264629

891	13518879 (1781, 1782)	Novel Protein sim. GBank gi 4959396 gb AAD3431.1 AF11248 - (AF112481) RAD54B protein [Homo sapiens]		UNCLASSIFIED	264636
892	87634157 (1783, 1784)	Novel Protein sim. GBank gi 545526 p bbs 43833 - LBP- 1b-transcription factor binding to initiation site of HIV-1 (alternatively spliced) [human, Namatawa cells, Peptide, 541 aa]		transcriptfactor	22278986, 22278999, 29331828, 35696052, 264808, 264809, 265009, 265011, 264602, 265019, 264766, 21906765, 21906766, 21908769, 265020, 285021, 56528486
893	78168037 (1785, 1786)	Novel Protein sim. GBank gi 2829688 sp P80608 CYSK_MAIZE - CYSTEINE SYNTHASE (O-ACETYL SERINE SULFHYDRYLASE) (O- ACETYL SERINE (THIOL) LYASE) (CSASE)		synthase	264689, 263967
894	11102240 (1787, 1788)				263978
895	80239668 (1789, 1790)			UNCLASSIFIED	264508, 264600, 264555, 264559
896	79747803 (1791, 1792)				264632
897	94991823 (1793, 1794)			Contains protein domain (PF00047) - Immunoglobulin domain	264686, 29331828, 264511
898	87695109 (1795, 1796)				56182575, 60432289, 56182435, 60432229, 55811957, 22279000, 264486
899	11100463 (1797, 1798)				264601
900	80499768 (1799, 1800)	Novel Protein sim. GBank gi 1750127 (U86480) - YncC [Bacillus subtilis]		transport	264769, 264691, 264563
901	80502410 (1801, 1802)	Novel Protein sim. GBank gi 3122879 sp O07438 SYA_MYCTU - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)		Contains protein domain (PF01411) - RNA synthetases class II (A)	264907, 264602, 264605, 264769, 35695917, 18108376, 264563
902	80503301 (1803, 1804)	Novel Protein sim. GBank gi 3355701 emb CAA20001  - (AL031124) 3-isopropylmalate dehydratase large subunit [Streptomyces coelicolor]		Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	264909, 265008, 264602, 264604, 264769, 264689, 264693
903	82060206 (1805, 1806)	Novel Protein sim. GBank gi 2960120 emb CAA18018.1  - (AL022121) gipK [Mycobacterium tuberculosis]		Contains protein domain (PF00370) - FGY family of carbohydrate kinases	35696052, 264905, 264510, 264511, 264512, 264605, 264760, 18108351, 264762, 264687, 264768, 264769, 264688, 21906764, 35695917, 27486282, 35695855, 264634, 264636, 264486
904	20451078 (1807, 1808)	Novel Protein sim. GBank gi 728887 sp P40906 ARGI - ARGINASE		Contains protein domain (PF00491) - Arginase family	264604
905	9338483 (1809, 1810)	Novel Protein sim. GBank gi 4567200 gb AAD23616.1 AC007168 - (AC007168) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264909
906	80052628 (1811, 1812)	Novel Protein sim. GBank gi 3738200 emb CAA21292  - (AL031855) putative vacuolar membrane protein [Schizosaccharomyces pombe]			264595, 264605
907	87913201 (1813, 1814)				
908	11754482 (1815, 1816)			UNCLASSIFIED	60432289, 264601, 264690
909	202727807 (1817, 1818)	Novel Protein sim. GBank gi 3668940 db BAA34296  - (AB015054) Alg2 [Rhizomucor pusillus]		UNCLASSIFIED	264638
910	16776206 (1819, 1820)	Novel Protein sim. GBank gi 4589726 db BAA76883.1  - (AB003137) DnaJ homolog protein [Salix gligiana]		UNCLASSIFIED	264602
				Contains protein domain (PF00684) - DnaJ central domain (4 repeats)	265009



911	87454340 (1821, 1822)	Novel Protein sim. GBank gi 548774 sp P35685 RL7A_ORYSA - 60S RIBOSOMAL PROTEIN L7A		ribosomalprot	265010, 264604, 60432113
912	20448863 (1823, 1824)	Novel Protein sim. GBank gi 2314008 gb AAD07921.1  - (AE000597) CDP-diglyceride hydrolase (cdh) [Helicobacter pylori 26895]		hydrolase	264559
913	20469357 (1825, 1826)	Novel Protein sim. GBank gi 417657 sp Q03604 RIR1_CAEEL - PROBABLE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (RIBONUCLEOTIDE REDUCTASE)	Contains protein domain (PF00317) - Ribonucleotide reductase	UNCLASSIFIED - reductase	264604 264636
914	79183351 (1827, 1828)	Novel Protein sim. GBank gi 568957 dbj BAA83069.1  - (AB028040) KIAA1117 protein [Homo sapiens]			18108398, 22278996, 66714117, 264906, 264591, 21908768, 265020, 55811576, 264638 264595
915	87606703 (1829, 1830)	Novel Protein sim. GBank gi 568957 dbj BAA83069.1  - (AB028040) KIAA1117 protein [Homo sapiens]		UNCLASSIFIED	264605
916	78444091 (1831, 1832)	Novel Protein sim. GBank gi 4186110 emb CAA71790  - (Y10831) putative integrase [Ralstonia eutropha]		UNCLASSIFIED	264259, 29331826, 264908, 265019, 264448, 265020, 264635, 83373044
917	20185985 (1833, 1834)	Novel Protein sim. GBank gi 1655699 emb CAA69032  - (Y07752) pherophorin-S [Volvox carter]		synthase	265006, 264512, 264600, 264602, 264604, 264768, 18108370, 264563
918	91228795 (1835, 1836)	Novel Protein sim. GBank gi 1655699 emb CAA69032  - (Y07752) pherophorin-S [Volvox carter]			264508
919	80436785 (1837, 1838)	Novel Protein sim. GBank gi 5689968 emb CAB52005.1  - (AL109663) putative membrane protein [Streptomyces coelicolor A3(2)]		peptidase	264600
920	78606095 (1839, 1840)	Novel Protein sim. GBank gi 1168448 sp Q05813 AMP1_STRLI - XAA-PRO AMINOPEPTIDASE I (X-PRO AMINOPEPTIDASE I) (AMINOPEPTIDASE P I) (APP) (PEPP I) (AMINOACYLPROLINE AMINOPEPTIDASE I)		UNCLASSIFIED	265019, 22279002 264907, 265008, 265009, 265010, 18108351, 264689, 265021, 18108370, 18108374, 18108385 264510 264603
921	19858634 (1841, 1842)	Novel Protein sim. GBank gi 3850084 emb CAA2191.1  - (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]			
922	78982605 (1843, 1844)	Novel Protein sim. GBank gi 267079 sp P29514 TB86_ARATH - TUBULIN BETA-6 CHAIN		tubulin	
923	8685830 (1845, 1846)	Novel Protein sim. GBank gi 267079 sp P29514 TB86_ARATH - TUBULIN BETA-6 CHAIN			
924	21431341 (1847, 1848)	Novel Protein sim. GBank gi 2497688 sp Q06863 PAFA_MOUSE - PLATELET- ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (PAF ACETYLHYDROLASE) (PAF 2- ACYLHYDROLASE) (LDL-ASSOCIATED PHOSPHOLIPASE A2) (LDL-PLA(2)) (2-ACETYL-1- ALKYLGLYCEROPHOSPHOCHOLINE ESTERASE) (1- ALKYL-2-ACETYLGLYCEROPHO...		esterase	
925	20630332 (1849, 1850)	Novel Protein sim. GBank gi 3882325 dbj BAA34522.1  - (AB018345) KIAA0802 protein [Homo sapiens]			55811957, 263972, 264639
926	78397657 (1851, 1852)	Novel Protein sim. GBank gi 3882325 dbj BAA34522.1  - (AB018345) KIAA0802 protein [Homo sapiens]			
927	37036201 (1853, 1854)			UNCLASSIFIED	264769

928	80070610 (1855, 1856)	Novel Protein sim. GBank gij156146 (M30316) - xanthine dehydrogenase (AA at 2538) [Caliphora vicina]		dehydrogenase	264605	
929	20630336 (1857, 1858)			UNCLASSIFIED	264603	
930	5496348 (1859, 1860)	Novel Protein sim. GBank gij4115936 [gbjAADO3446.1] - (AF118223) No definition line found [Arabidopsis thaliana]			264259	
931	10245731 (1861, 1862)	Novel Protein sim. GBank gij4490809 [embjCAB38642.1] - (AJ133495) ribonucleotide reductase major subunit [Staphylococcus aureus]	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	264486	
932	80420613 (1863, 1864)	Novel Protein sim. GBank gij5459396 [embjCAB50754.1] - (AL098839) putative integral membrane transport protein [Streptomyces coelicolor]		transport	264508, 264905, 264906, 264909, 264600, 264602, 264603, 264605, 264766, 264558, 18108387, 264486	
933	94326010 (1865, 1866)	Novel Protein sim. GBank gij5689523 [dbjBAA83045.1] - (AB029016) KIAA1093 protein [Homo sapiens]	Contains protein domain (PF00827) - UBA domain	UNCLASSIFIED	264508, 264686, 264693, 27486261, 18108370, 65274791, 264636, 264559, 22279002	
934	80039105 (1867, 1868)	Novel Protein sim. GBank gij119111 [spjP12978] [EBN2_EBV - EBNA-2 NUCLEAR PROTEIN]			264369, 22279002	
935	80063162 (1869, 1870)			UNCLASSIFIED	264604, 264605, 264693, 18108370, 18108374	
936	80026632 (1871, 1872)	Novel Protein sim. GBank gij845686 (M32103) - ORF-27 [Staphylococcus aureus]		UNCLASSIFIED	22278995, 22278996, 264602, 264687, 32833986, 18108387	
937	80250273 (1873, 1874)	Novel Protein sim. GBank gij1360669 [pirjCGHUV - collagen alpha 1(V) chain precursor - human]		synthase	263978	
938	80026633 (1875, 1876)	Novel Protein sim. GBank gij2226243 [embjCAA74531.1] - (Y14083) hypothetical protein [Bacillus subtilis]		UNCLASSIFIED	264602	
939	11071694 (1877, 1878)				264600	
940	94144252 (1879, 1880)	Novel Protein sim. GBank gij3560166 [embjCAA20678] - (AL031525) ubiquitin carboxyl-terminal hydrolase [Schizosaccharomyces pombe]	Contains protein domain (PF00442) - Ubiquitin carboxyl-terminal hydrolases family 2	ubiquitin	264905, 264906, 264907, 264908, 264909, 264511, 264910, 264592, 33657402, 264596, 264758, 264760, 264683, 264766, 264768, 264769, 33657023, 33657109, 264628, 264629, 264630, 264635, 264636, 264555, 264637, 264556, 264638, 264639, 83373044, 18108385, 264565, 18108391	
941	11398414 (1881, 1882)				264593	
942	19484122 (1883, 1884)			UNCLASSIFIED	264760	
943	80080258 (1885, 1886)	Novel Protein sim. GBank gij4033729 (AF038595) - apolipoprotein N-acyltransferase [Pseudomonas aeruginosa]		UNCLASSIFIED	264600, 264687, 264689, 264563	
944	80216086 (1887, 1888)	Novel Protein sim. GBank gij2494764 [spjQ50729] [GUAA_MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)]		synthase	264511, 264603	
945	80052477 (1889, 1890)	Novel Protein sim. GBank gij732353 [spjP39606] [WCH_BACSU - HYPOTHETICAL 36.6 KD PROTEIN IN QOXD-VPR INTERGENIC REGION]	Contains protein domain (PF00818) - Ice nucleation protein repeat	UNCLASSIFIED	264906, 264604, 264605, 265020, 18108387	
946	78248402 (1891, 1892)				265017	

947	81802698 (1893, 1894)	Novel Protein sim. GBank gij2896770[embjCAA17247] - (AL021899) hypothetical protein Rv2033c [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - Inositol monophosphatase family	phosphatase	18108394, 22278996, 264907, 264909, 265008, 265009, 264910, 264758, 264600, 264602, 265018, 264605, 264769, 264689, 264693
948	88165538 (1895, 1896)	Novel Protein sim. GBank gij2827284 (AF041037) - novel antagonist of EGF signaling [Homo sapiens]		lgr	18108398, 56182575, 22278997, 22278999, 60432049, 29331822, 29331826, 264907, 56182435, 55811386, 265011, 264600, 265017, 265018, 265019, 18108351, 265020, 265021, 265022, 27486265, 263972, 55811576, 264638, 80170394, 264566
949	88081786 (1897, 1898)	Novel Protein sim. GBank gij4507985[refjNP_003427.1]pZNF1 - zinc finger protein 135 (clone pHZ-17)	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	29331825, 21906764, 27486261, 21906766, 52644298, 33657349, 87168518, 56994075, 265020, 265021, 87168559, 52644150, 264637
950	78485872 (1899, 1900)	Novel Protein sim. GBank gij1079461[prj]S43865 - Cytochrome b, type II - potoro (fragment)	Contains protein domain (PF00038) - Intermediate filament proteins	struct	264683, 18108361
951	20451411 (1901, 1902)	Novel Protein sim. GBank gij5420387[embjCAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264604
952	79566954 (1903, 1904)	Novel Protein sim. GBank gij5305702[gbjAAD41779.1]AF12686 - (AF126867) calpain-like protease [Mus musculus]		cathepsin	264910, 264691
953	10186003 (1905, 1906)	Novel Protein sim. GBank gij2495642[spjQ47142]VFHS, ECOLI - HYPOTHETICAL 41.6 KD PROTEIN IN CSIE-GLYA INTERGENIC REGION		transport	264510
954	9893328 (1907, 1908)	Novel Protein sim. GBank gij2360965 (AF016253) - D-amino acid dehydrogenase [Klebsiella aerogenes]		dehydrogenase	264508
955	95313410 (1909, 1910)	Novel Protein sim. GBank gij5454064[refjNP_006318.1]pSIPI - SYT interacting protein	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	56994075, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264910, 264758, 264759, 265010, 264601, 264760, 18108351, 264762, 264763, 264764, 264766, 264686, 264767, 264687, 264768, 264769, 264689, 264628, 264629, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 55182323, 264639, 18108388, 264563, 264564
956	80064224 (1911, 1912)	Novel Protein sim. GBank gij2052129[embjCAB08155] - (Z94752) rimJ [Mycobacterium tuberculosis]			264605
957	80056206 (1913, 1914)			UNCLASSIFIED	264603, 18108362
958	80036446 (1915, 1916)	Novel Protein sim. GBank gij1709787[spjQ00451]PRF1_LYCES - 36.4 KD PROLINE-RICH PROTEIN		UNCLASSIFIED	264908, 264910, 264762, 263978, 264637
959	80026647 (1917, 1918)	Novel Protein sim. GBank gij2131050[embjCAB09260] - (Z95844) opaA [Mycobacterium tuberculosis]		UNCLASSIFIED	264602, 264692
960	37815406 (1919, 1920)	Novel Protein sim. GBank gij2129478[prj]S51939 - chitinase (EC 3.2.1.14) precursor - beet		UNCLASSIFIED	264259
961	20567383 (1921, 1922)				263978
962	11399318 (1923, 1924)				264593

963	80590374 (1925, 1926)			UNCLASSIFIED	264510, 264288, 264555, 264556, 264559, 264486
964	79832019 (1927, 1928)	Novel Protein sim. GBank gij4589622[jdb][BAA76833.1] - (AB023206) KIAA0989 protein [Homo sapiens]		UNCLASSIFIED	264112, 264910, 264689
965	91229485 (1929, 1930)	Novel Protein sim. GBank gij5420387[emb][CAB46670.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264488, 265017, 264448, 264634, 264558, 83373044
966	95292815 (1931, 1932)			UNCLASSIFIED	264906, 264592, 264596, 264604, 264768, 21906764, 264692, 264693, 264629, 264636, 264638
967	79255708 (1933, 1934)	Novel Protein sim. GBank gij1731207[sp][Q11156][RGX3 - MYCTU - SENSORY TRANSDUCTION PROTEIN REGX3	(Contains protein domain (PF00072) - Response regulator receiver domain	- phosphatase	264760
968	78560269 (1935, 1936)	Novel Protein sim. GBank gij2661836[emb][CAA75187] - (Y14964) putative transport protein [Methylophilus methylotrophus]		transport	264693
969	79918470 (1937, 1938)	Novel Protein sim. GBank gij5419878[emb][CAB46422.1] - (AL096747) hypothetical protein [Homo sapiens]	(Contains protein domain (PF00096) - Zinc finger, C2H2 type	- dna_ma_bind	35696286, 264685, 264686, 35695917, 264692, 18108374, 264635
970	95085947 (1939, 1940)			UNCLASSIFIED	18108392, 18108394, 18108398, 22278995, 22278996, 22278998, 22278999, 29147620, 264828, 265006, 265007, 265008, 265009, 18108348, 33109854, 265010, 265011, 18108351, 264288, 21906767, 21906768, 18108370, 18108374, 18108377, 264630, 264635, 18108380, 83373044, 18108387, 18108388
971	78919770 (1941, 1942)			UNCLASSIFIED	265007, 265020, 22279002
972	20710704 (1943, 1944)				264557
973	20370183 (1945, 1946)	Novel Protein sim. GBank gij1723119[sp][P53990][Y174_HUMAN - HYPOTHETICAL PROTEIN KIAA0174			264604
974	80057103 (1947, 1948)			UNCLASSIFIED	264565
975	10196018 (1949, 1950)			UNCLASSIFIED	264510
976	80205742 (1951, 1952)	Novel Protein sim. GBank gij3881459[emb][CAA92988.1] - (Z58753) predicted using GeneFinder. Similarity to Yeast hypothetical protein YIK9 (SW:YIK9_YEAST); cDNA EST EMBL:D27680 comes from this gene; cDNA EST EMBL:D27679 comes from this gene; cDNA EST EMBL:D84477 comes from this ge...		UNCLASSIFIED	264508, 264906, 264758, 264632, 264639, 264563
977	10355349 (1953, 1954)	Novel Protein sim. GBank gij549456[sp][Q05335][XYS3_PSEPU - XYLDLEGF OPERON TRANSCRIPTIONAL ACTIVATOR 3		UNCLASSIFIED	264906
978	80025927 (1955, 1956)			UNCLASSIFIED	264600, 264602, 264603, 264604
979	80447820 (1957, 1958)	Novel Protein sim. GBank gij3171904[emb][CAA75669] - (Y15908) DIA-12C protein [Homo sapiens]		UNCLASSIFIED	264767, 264768, 265006, 265007, 264906
980	80025928 (1959, 1960)			UNCLASSIFIED	264600, 264602, 264605
981	80098550 (1961, 1962)	Novel Protein sim. GBank gij3599940 (AF017368) - fadogenital dysplasia protein 2 [Mus musculus]		UNCLASSIFIED	264692, 264555, 264556, 264557, 264559

982	80195670 (1963, 1964)	Novel Protein sim. GBank gij2950220[emb CAA71575] - (Y10545) fused-cdb [Escherichia coli]		UNCLASSIFIED	264404
983	90995041 (1965, 1966)	Novel Protein sim. GBank gij476389[prj B43402 - myosin heavy chain-B, neuronal - chicken]		struct	65274572, 56182575, 264908, 264909, 265007, 265008, 264758, 265010, 55811150, 33657023, 264634, 264557, 264558
984	20466876 (1967, 1968)	Novel Protein sim. GBank gij3451504[emb CAA07660.1] - (A1007747) hypothetical protein BbLPS1.21 [Bordetella bronchiseptica]	Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	264605
985	65461368 (1969, 1970)			transferase	56182435, 264600
986	87102888 (1971, 1972)			UNCLASSIFIED	264106, 264110, 265020, 60170615
987	79867231 (1973, 1974)				264909
988	19858661 (1975, 1976)			UNCLASSIFIED	264600
989	88095329 (1977, 1978)			UNCLASSIFIED	264508, 265017, 264534, 264564
990	88057746 (1979, 1980)	Novel Protein sim. GBank gij5725508[gb AAD48080.1 AF060152] MIEH1 protein [Homo sapiens]	Contains protein domain (PF01421) - Reprolysin (M12B) family zinc metalloprotease	oxidase	264259, 264908, 265009, 264910, 264596, 264369, 264288, 264766, 264628, 264635, 264566
991	10106140 (1981, 1982)			UNCLASSIFIED	264909
992	79845694 (1983, 1984)	Novel Protein sim. GBank gij2105049[emb CAB08835] - (Z95436) hypothetical protein Rv3645 [Mycobacterium tuberculosis]	Contains protein domain (PF00211) - Adenylate and Guanylate cyclase catalytic domain	UNCLASSIFIED	264508, 264593
993	10814053 (1985, 1986)				
994	11090590 (1987, 1988)	Novel Protein sim. GBank gij3329297 (AE001355) - Ribonucleoside Reductase, Large Chain [Chlamydia trachomatis]		reductase	264907
995	94321611 (1989, 1990)	Novel Protein sim. GBank gij5106572[gb AAD39760.1 AF14394] - transcriptional activator SRCAP [Homo sapiens]	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	helicase	18108398, 65274572, 22278986, 264490, 60432049, 28331827, 28146498, 264508, 264905, 264907, 264908, 56182435, 265008, 264591, 264592, 60432229, 60431735, 33657402, 264595, 264758, 21906754, 265010, 265017, 265018, 264605, 264760, 264448, 264763, 284768, 21906765, 21908768, 21906789, 55811957, 264692, 264693, 264629, 35696423, 55811576, 35695855, 264636, 264555, 264556, 264558, 83373044, 22279002, 264563
996	91013745 (1991, 1992)	Novel Protein sim. GBank gij2911719 (AC004227) - KIA001LB [Homo sapiens]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF)	phosphatase	83373044, 29331824, 263878, 55811957, 56526486, 87168518, 264910, 264906, 264565, 264566, 264693, 264766
997	80503347 (1993, 1994)	Novel Protein sim. GBank gij2649101 (AE001001) - ABC transporter, ATP-binding protein [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - ABC transporter	transport	35696286, 22278997, 22278999, 264508, 264905, 264908, 265010, 264600, 264602, 264605, 264688, 264769, 265021, 264565, 18108391
998	11397390 (1995, 1996)	Novel Protein sim. GBank gij123530[sp P04929 HRPX_PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR]			264595

999	11768047 (1997, 1998)	Novel Protein sim. GBank gij2506697[sp P46490]YFCA_HAEIN - HYPOTHETICAL PROTEIN H10198		UNCLASSIFIED	264682
1000	20727944 (1999, 2000)			UNCLASSIFIED	264602
1001	86673131 (2001, 2002)	Novel Protein sim. GBank gij2224689[dbj BAA20833] - (AB002377) KIAA0379 [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	kinase	60432049, 264907, 264909, 264511, 264603, 264683, 264684, 264687, 264689, 29148627, 21906769, 264692, 18108385, 22279000
1002	80189603 (2003, 2004)	Novel Protein sim. GBank gij586121[sp P37709]TRHY_RABIT - TRICHOHYALIN		struc	265009, 264369, 265020
1003	17933491 (2005, 2006)				265019
1004	16314987 (2007, 2008)	Novel Protein sim. GBank gij854065[embj CAA58337] - (X83413) U88 [Human herpesvirus 6]			264635
1005	79617144 (2008, 2010)	Novel Protein sim. GBank gij114073[sp P07672]APT_ECOLI - ADENINE PHOSPHORIBOSYLTRANSFERASE (APRT)		UNCLASSIFIED	264508
1006	37815429 (2011, 2012)			UNCLASSIFIED	264259
1007	79620871 (2013, 2014)	Novel Protein sim. GBank gij4062979[dbj BAA36210.1] - (AB017138) epsilon subunit of malonate decarboxylase [Pseudomonas putida]		synthase	264905
1008	88094444 (2015, 2016)	Novel Protein sim. GBank gij2808807[embj CAA04607.1] - (AJ001206) putative trehalose synthase [Streptomyces coelicolor]		synthase	265007, 264602, 264605, 264760, 264636
1009	57451289 (2017, 2018)	Novel Protein sim. GBank gij3639077 (AF090113) - AMPA receptor binding protein [Rattus norvegicus]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF)	kinase	264102, 264288
1010	94672537 (2019, 2020)	Novel Protein sim. GBank gij3746332 (AF016307) - possible NADH-dependent oxidase, may function as a demethylase [Sinorhizobium meliloti]		dehydrogenase	264592
1011	85546916 (2021, 2022)	Novel Protein sim. GBank gij2342847[gb AAB86591.1] - (U90653) DHC-domain-containing cysteine-rich protein [Homo sapiens]		UNCLASSIFIED	35696052, 264905, 264784, 264768, 35695917, 264629
1012	95294456 (2023, 2024)	Novel Protein sim. GBank gij3413411[embj CAA20272] - (AL031231) guanosine pentaphosphate synthetase/ polyribonucleotide nucleotidyltransferase [Streptomyces coelicolor]	Contains protein domain (PF00013) - KH domain	phosphorylase	35696052, 264905, 264600, 264601, 264602, 264605, 264762, 264768, 264768, 264689
1013	86095772 (2025, 2026)			UNCLASSIFIED	264591, 21906768
1014	86608828 (2027, 2028)				29331824, 265019, 265020

1015	95418879 (2028, 2030)	Novel Protein sim. GBank gij4159995 (AF063095) - SELIL [Mus musculus]	Contains protein domain (PF00040) - Fibronectin type II domain	struct	22278994, 22278995, 56994075, 22278996, 22278999, 264259, 29331825, 29331828, 264907, 56182435, 264510, 264591, 264593, 60433356, 264594, 55812038, 264758, 21906754, 33657084, 265010, 264600, 265017, 265018, 265019, 18108351, 21906765, 21906768, 21906767, 21906768, 55811957, 265022, 33657023, 65274620, 33657182, 32833986, 18108370, 18108377, 55811576, 35696423, 264630, 22279000, 264565
1016	79559694 (2031, 2032)	Novel Protein sim. GBank gij250969spjP41407JACPD, ECOLI - ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)		esterase	264686, 264693
1017	11069213 (2033, 2034)	Novel Protein sim. GBank gij5103943jdbjBAA79259.1] - (AP000059) 802aa long hypothetical oligopeptide-binding protein oppA [Aeropyrum pernix]	Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5	transport	264600
1018	80072430 (2035, 2036)	Novel Protein sim. GBank gij4493973jembjCAB39032.1] - (AL034559) predicted using hexExon; MAL3P7.14 (PFC0925w), Hypothetical protein, len: 489 aa [Plasmodium falciparum]			22278996, 29148627, 264563
1019	11703807 (2037, 2038)			UNCLASSIFIED	264686
1020	80234432 (2039, 2040)				264508, 264509, 264512, 264600, 264762, 264769, 264689, 18108370, 264636, 264638, 264486
1021	37036243 (2041, 2042)	Novel Protein sim. GBank gij4633807jgbjAAD26859.1jAF12779 - (AF127795) trehalose biosynthetic enzyme TreY [Rhizobium leguminosarum bv. viciae]		synthase	264769
1022	80502627 (2043, 2044)	Novel Protein sim. GBank gij1781230jembjCAB06277] - (Z83867) hypothetical protein Rv3137 [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - Inositol monophosphatase family	phosphatase	35696052, 264508, 265008, 265009, 264769, 18108387, 264563
1023	11399341 (2045, 2046)	Novel Protein sim. GBank gij3777495 (U92083) - calcium transporting ATPase [Pichia angusta]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase-associated	264593
1024	80057129 (2047, 2048)			UNCLASSIFIED	52646842, 33657402, 33657023, 18108379, 55811576, 264631, 264556, 264557, 264559, 18108388, 264566
1025	79644200 (2049, 2050)	Novel Protein sim. GBank gij3483045jembjCAA20556] - (AL031371) putative transport system permease protein [Streptomyces coelicolor]		transport	264693
1026	80025946 (2051, 2052)	Novel Protein sim. GBank gij1174922spjQ02322jUVRD_HAEIN - DNA HELICASE II		helicase	264602
1027	117659234 (2053, 2054)	Novel Protein sim. GBank gij475728refjNP_004886.1jPAGTA - angiotensin/vasopressin receptor AII/AVP-like		UNCLASSIFIED	265017

1028	20297928 (2055, 2056)	Novel Protein sim. GBank gij2791409[emb]CAA16003] - (AL021184) acn [Mycobacterium tuberculosis]	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	UNCLASSIFIED	264600
1029	94665090 (2057, 2058)			UNCLASSIFIED	264595
1030	86095343 (2059, 2060)			UNCLASSIFIED	264907, 264908, 264510, 264512, 265009, 265010, 265011, 264600, 264602, 264603, 264605, 264769, 18108372, 18108374
1031	95289117 (2061, 2062)			UNCLASSIFIED	264905, 264906, 264909, 264595, 264692, 264630, 264634, 264638
1032	94673275 (2063, 2064)	Novel Protein sim. GBank gij4503895[ref]NP_000145.1[pGALK - galactokinase 1		UNCLASSIFIED	264689
1033	86464818 (2065, 2066)	Novel Protein sim. GBank gij2982990 [AE000682] - hypothetical protein [Aquifex aeolicus]		UNCLASSIFIED	35896052, 264908, 264510, 18108354, 264687, 264769, 264689, 80431602, 18108385, 264486
1034	78245937 (2067, 2068)	Novel Protein sim. GBank gij405895 [U00007] - methionyl-tRNA synthetase [Escherichia coli]		UNCLASSIFIED	264906
1035	78956355 (2069, 2070)			UNCLASSIFIED	264692
1036	85804988 (2071, 2072)			UNCLASSIFIED	264905, 66712502, 264908, 264768
1037	87896058 (2073, 2074)			UNCLASSIFIED	29331824, 264909, 60433438, 265019
1038	20481015 (2075, 2076)			UNCLASSIFIED	264604, 264634
1039	87260021 (2077, 2078)	Novel Protein sim. GBank gij790819 (L39891) - polycystic kidney disease-associated protein [Homo sapiens]	Contains protein domain (PF01477) - PLAT/LH2 domain	UNCLASSIFIED	264092, 264093, 264094, 264683, 264689, 263967
1040	80026840 (2078, 2080)	Novel Protein sim. GBank gij2352095 (U97022) - DNA topoisomerase 1 [Fervidobacterium islandicum]	Contains protein domain (PF01131) - Prokaryotic DNA topoisomerase	isomerase	264595
1041	10156682 (2081, 2082)	Novel Protein sim. GBank gij3256535[dbj]BAA29218.1] - (AP0000001) 301aa long hypothetical 2-phosphoglycerate kinase [Pyrococcus horikoshii]		kinase	264907
1042	11084375 (2083, 2084)	Novel Protein sim. GBank gij2058299[emb]CAA66953] - (X98309) ARI protein [Drosophila melanogaster]			264605
1043	80057136 (2085, 2086)	Novel Protein sim. GBank gij1870167[emb]CAA70125] - (Y08921) msK [Streptomyces reticuli]	Contains protein domain (PF00005) - ABC transporter	transport	264565, 264567
1044	80025952 (2087, 2088)			UNCLASSIFIED	265006, 264602, 265017
1045	52415482 (2089, 2090)	Novel Protein sim. GBank gij5689890[emb]CAB52053.1] - (AL109732) hypothetical protein [Streptomyces coelicolor A3(2)]			29331825, 264637
1046	11754862 (2091, 2092)	Novel Protein sim. GBank gij854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 6]		helicase	264686
1047	37036258 (2093, 2094)	Novel Protein sim. GBank gij4210471[dbj]BAA74535.1] - (AB019033) orfSA [Pseudomonas sp.]		UNCLASSIFIED	264769
1048	79186400 (2095, 2096)	Novel Protein sim. GBank gij3413419[emb]CAA20279] - (AL031232) hypothetical protein SC10H5.07 [Streptomyces coelicolor]	Contains protein domain (PF00220) - Neurohypophysial hormones, N-terminal Domain	UNCLASSIFIED	264687
1049	81755108 (2097, 2098)	Novel Protein sim. GBank gij5051636[gb]AAD38326.1[AF07372] - (AF073727) EH domain-binding mitotic phosphoprotein [Homo sapiens]		UNCLASSIFIED	264905, 264634
1050	79471521 (2099, 2100)			UNCLASSIFIED	264686



1051	80475471 (2101, 2102)				UNCLASSIFIED	18108374, 264769, 265010, 265011, 264601, 265009, 264604, 264605, 264636, 18108351, 264692
1052	82442962 (2103, 2104)	Novel Protein sim. GBank gi 3123275 sp P35136 SERA_BACSU - D-3- PHOSPHOGLYCERATE DEHYDROGENASE (PGDH)	Contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases		dehydrogenase	264508, 264762, 264687, 264486
1053	94851640 (2105, 2106)	Novel Protein sim. GBank gi 5441319 emb CAB46717.1  - (AL034396) dJ1158B12.1 (zinc finger, X-linked, duplicated A) [Homo sapiens]				264686, 18108374, 29331824, 83373044, 21906754, 52645156, 56182435, 264689, 29331827, 27486261, 35696052, 21908765, 35696423, 21908768, 56182575, 21908769, 55811957, 87168518, 35696286, 22278997, 265020, 265011, 265021, 265022, 265007, 265018, 22279000, 22279002, 264482, 264906, 52644150, 264909, 264288, 29331822, 52645080, 264766
1054	79580225 (2107, 2108)				UNCLASSIFIED	264686
1055	80594138 (2109, 2110)	Novel Protein sim. GBank gi 5052508 gb AAD38584.1 AF145609) BcDNA, GH02833 [Drosophila melanogaster]	Contains protein domain (PF00270) - DEAD/DEAH box helicase		helicase	264907, 264602, 264681, 264288, 21908768, 33657109, 55810764, 35695855, 264631
1056	17882319 (2111, 2112)	Novel Protein sim. GBank gi 3021676 dbj BAA25358  - (D86033) RNA polymerase sigma-70 factor [Pseudomonas fluorescens]			mapolymerase	264906
1057	85687216 (2113, 2114)	Novel Protein sim. GBank gi 1226281 U50308) - No definition line found [Caenorhabditis elegans]			UNCLASSIFIED	264682
1058	80376578 (2115, 2116)					264764
1059	94662754 (2117, 2118)	Novel Protein sim. GBank gi 1170016 sp P46808 GREA_MYCLE - TRANSCRIPTION ELONGATION FACTOR GREA (TRANSCRIPT CLEAVAGE FACTOR GREA)			transcriptfactor	35696052, 35695855, 265009, 264636
1060	79481169 (2119, 2120)	Novel Protein sim. GBank gi 2499087 sp Q09332 UGGG_DROME - UDP- GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT)			glycoprotein	29146499, 264681, 264683, 264687
1061	11034025 (2121, 2122)	Novel Protein sim. GBank gi 90254 pir J28334 - protein- tyrosine-phosphatase (EC 3.1.3.48) Ly-5 precursor (B-cell variant) - mouse			phosphatase	264634
1062	39567937 (2123, 2124)	Novel Protein sim. GBank gi 3334200 sp O49954 GCSP_SOL_TU - GLYCINE DEHYDROGENASE (DECARBOXYLATING) PRECURSOR (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN)			dehydrogenase	264593
1063	8490481 (2125, 2126)	Novel Protein sim. GBank gi 2499986 sp Q41228 PSE1_NICSY - PHOTOSYSTEM I REACTION CENTRE SUBUNIT IV A PRECURSOR (PSI-E A)				264508
1064	78891763 (2127, 2128)	Novel Protein sim. GBank gi 82654 pir JUA0086 - 10K zein precursor - maize				265007, 265008, 18108351, 18108385

1065	80021208 (2129, 2130)	Novel Protein sim. GBank gij2120988[pri][S70682 - glycosyltransferase homolog - Bordetella pertussis]		transferase	264600, 264602, 264689
1066	17698879 (2131, 2132)	Novel Protein sim. GBank gij2506382[sp]P15042[DNLI_ECOLI - DNA LIGASE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (NAD+))]		synthase	265009
1067	10132178 (2133, 2134)				264909
1068	82062057 (2135, 2136)	Novel Protein sim. GBank gij4007669[emb][CAA22355] - (AL0344443) putative oxidoreductase [Streptomyces coelicolor]	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	264688, 18108362, 264558, 264600, 264760
1069	83002954 (2137, 2138)	Novel Protein sim. GBank gij4589484[dbj][BAA76770.1] - (AB023143) KIAA0928 protein [Homo sapiens]		UNCLASSIFIED	264604
1070	82101992 (2139, 2140)	Novel Protein sim. GBank gij120304[sp]P15932[FLGK_SALTY - FLAGELLAR HOOK-ASSOCIATED PROTEIN 1 (HAP1)]		UNCLASSIFIED	264604, 264760
1071	20710589 (2141, 2142)	Novel Protein sim. GBank gij1750397 (U81261) - glutamate synthase large subunit [Pseudomonas aeruginosa]		synthase	264602
1072	82356540 (2143, 2144)			UNCLASSIFIED	264687, 264688, 21906764, 35696052, 35695917, 35695855, 264600, 264601, 264602, 265009, 264605, 264508, 264905, 264690, 264906, 264762, 264628, 264768
1073	79814400 (2145, 2146)				264909
1074	80105982 (2147, 2148)	Novel Protein sim. GBank gij477532[pri][A49175 - Notch B protein - mouse (fragment)]	Contains protein domain (PF00008) - EGF-like domain	synthase	264906
1075	81850293 (2149, 2150)	Novel Protein sim. GBank gij3893109[emb][CAA76940] - (Y17920) CALO protein [Drosophila melanogaster]		UNCLASSIFIED	56994075, 22278998, 264594, 264757, 264596, 265018, 265019, 264681, 264369, 264688, 265020, 18108364, 18108374
1076	80477284 (2151, 2152)	Novel Protein sim. GBank gij1176203[sp]P46442[YHGM_ECOLI - HYPOTHETICAL 43.1 KD PROTEIN IN RPLM-HHOA INTERGENIC REGION (F375)]		ATPase-associated	264769
1077	79831334 (2153, 2154)	Novel Protein sim. GBank gij4033487[sp]Q44472[ITUD4_AGRVI - PUTATIVE HYDROXYPYRUVATE REDUCTASE]		kinase	264905
1078	20288874 (2155, 2156)			UNCLASSIFIED	264600
1079	80494518 (2157, 2158)	Novel Protein sim. GBank gij3413828[emb][CAA20296] - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	18108394, 264769, 264634, 264636
1080	11767188 (2159, 2160)			UNCLASSIFIED	264684
1081	94747080 (2161, 2162)				83373044, 265019, 22279002, 264482, 18108351, 264682, 264908, 264693, 264487
1082	81490656 (2163, 2164)			UNCLASSIFIED	264758, 264768, 264769, 21906767, 264511, 264910, 264634, 264635, 264905, 264636, 264906, 264637, 264907, 264908, 264764, 264638, 20281099, 264766, 264595

1083	87446717 (2165, 2166)	Novel Protein sim. GBank gij1728945 sp Q10523 Y01N_MYCTU - HYPOTHETICAL 44.6 KD PROTEIN CY427.23		UNCLASSIFIED	60424179, 264905, 264906, 264510, 60432229, 264759, 87168474, 264605, 264769, 264689, 18108364, 18108376, 35695855, 264636
1084	37789308 (2167, 2168)	Novel Protein sim. GBank gij18384 sp P32057 WCAI_ECOLI - PUTATIVE COLANIC ACID BIOSYNTHESIS GLYCOSYL TRANSFERASE WCAI		UNCLASSIFIED	264769
1085	86475368 (2169, 2170)	Novel Protein sim. GBank shock protein 60 [Tsukamurella tyrosinohens]	Contains protein domain (PF00118) - eph		60432229, 264687
1086	79608269 (2171, 2172)	Novel Protein sim. GBank gij1172958 sp P46176 RL14_ACYKS - 50S RIBOSOMAL PROTEIN L14	Contains protein domain (PF00236) - Ribosomal protein L14	ribosomalprot	264486
1087	78603979 (2173, 2174)	Novel Protein sim. GBank gij4160198 emb CAA15431  - (AL008583) dJ327J16.3 (novel CHROMObox family protein) [Homo sapiens]	Contains protein domain (PF00385) - 'chromo' (CHROMatin Organization Modifier) domain	helicase	29331827, 264693
1088	79854963 (2175, 2176)	Novel Protein sim. GBank phosphoglucosyltransferase/phosphomannomutase [Aquifex aerophilus]		UNCLASSIFIED	264905, 264601, 18108387
1089	80218800 (2177, 2178)	Novel Protein sim. GBank gij4981768 gb AAD36280.1 AE001778 - (AE001778) NADH dehydrogenase, 30 kDa subunit, putative [Thermotoga maritima]	Contains protein domain (PF00329) - Respiratory-chain NADH dehydrogenase, 30 kDa subunit	UNCLASSIFIED	264488, 264511, 265011, 264682, 264768, 264689, 21908784, 35695917, 285020, 32833986, 18108370, 35695855
1090	11083825 (2179, 2180)	Novel Protein sim. GBank gij4007680 emb CAA22366  - (AL034443) putative oxidoreductase [Streptomyces coelicolor]			264604
1091	12817471 (2181, 2182)	Novel Protein sim. GBank gij2495562 sp P77239 YLCD_ECOLI - HYPOTHETICAL 44.3 KD PROTEIN IN NFRB-PHEP INTERGENIC REGION PRECURSOR		UNCLASSIFIED	264637
1092	80252286 (2183, 2184)	Novel Protein sim. GBank gij2060098 emb CAA17996.1  - (AL022121) ntlh [Mycobacterium tuberculosis]	Contains protein domain (PF00730) - Endonuclease III	nuclease	264566 264769, 35695917, 35695855, 264600, 264602, 264603, 264605, 18108351
1093	80496304 (2185, 2186)	Novel Protein sim. GBank gij1001642 dbj BAA10373  - (D84002) dGTP triphosphohydrolase [Synecococcus sp.]		UNCLASSIFIED	264686
1094	10880972 (2187, 2188)	Novel Protein sim. GBank gij458558 emb CAB40855.1  - (AL049628) putative adenine glycosylase [Streptomyces coelicolor]	Contains protein domain (PF00455) - Bacterial regulatory proteins, deoR family	nuclease	264906, 265007, 264595, 264600, 264602, 264603, 264604, 264605, 264762, 264768, 264769, 264636, 264558, 18108387, 60432113, 264482, 264486
1095	87457250 (2189, 2190)	Novel Protein sim. GBank gij115001 sp P19208 BIOB_BACSH - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	264600, 264602, 264603, 264604, 264605, 35695917, 264692, 264631
1096	80025977 (2191, 2192)	Novel Protein sim. GBank gij114135 sp P08205 JARGA_ECOLI - AMINO-ACID ACETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS)		UNCLASSIFIED	265019 264687
1097	79239560 (2193, 2194)	Novel Protein sim. GBank		synthase	
1098	79186424 (2195, 2196)	Novel Protein sim. GBank			

1099	39523838 (2197, 2198)	Novel Protein sim. GBank gi 3915144 sp O33017 TRMD_MYCLE - TRNA (GUANINE- NT)-METHYLTRANSFERASE (M1G- METHYLTRANSFERASE) (TRNA [GM37] METHYLTRANSFERASE)		UNCLASSIFIED	264603
1100	85736571 (2189, 2200)	Novel Protein sim. GBank gi 3023255 sp Q64420 ACOD_MESAU - ACYL-COA DESATURASE (STEAROYL-COA DESATURASE) (FATTY ACID DESATURASE) (DELTA(9)-DESATURASE)		desaturase	264259, 264636
1101	80491857 (2201, 2202)	Novel Protein sim. GBank gi 1174735 sp P43012 TOP1_HAEIN - DNA TOPOISOMERASE I (OMEGA-PROTEIN) (RELAXING ENZYME) (UNITWISTING ENZYME) (SWIVELASE)	Contains protein domain (PF01396) - Topoisomerase DNA binding C4 zinc finger	isomerase	264769
1102	79777614 (2203, 2204)			UNCLASSIFIED	264910, 264909
1103	81897259 (2205, 2208)	[Drosophila melanogaster] Novel Protein sim. GBank gi 1806596 (U81788) - kinesin-73		struct	264757
1104	95003115 (2207, 2208)	Novel Protein sim. GBank gi 2935448 (AF048976) - synaptic ras GTPase-activating protein p135 Syngap [Rattus norvegicus]		UNCLASSIFIED	29331822, 21906754, 264555, 264556, 264558, 22279002
1105	80255121 (2209, 2210)			UNCLASSIFIED	264566
1106	79314110 (2211, 2212)			UNCLASSIFIED	264555, 264369
1107	80470018 (2213, 2214)				264908, 264769
1108	80440816 (2215, 2216)	Novel Protein sim. GBank gi 1173421 sp P43416 SECY_STRSC - PREPROTEIN TRANSLUCASE SECY SUBUNIT		transport	264907, 264510, 264511, 264600, 264602, 264605, 264768, 264769
1109	80064615 (2217, 2218)	Novel Protein sim. GBank gi 2995310 emb CAA18338  - (AL022268) putative ATP-dependent helicase [Streptomyces coelicolor]		helicase	264602, 264605, 264636
1110	80503554 (2219, 2220)				
1111	80071744 (2221, 2222)	Novel Protein sim. GBank gi 2622039 (AE000868) - type I restriction modification system, subunit S [Methanobacterium thermoautotrophicum]			264908, 264593, 265010, 264601, 264603, 264604, 264605, 264682, 264769, 264693, 264636 18108370, 264557
1112	95010088 (2223, 2224)				
1113	82456352 (2225, 2226)	Novel Protein sim. GBank gi 3218376 emb CAA18628  - (AL023862) putative oxidoreductase [Streptomyces coelicolor]		UNCLASSIFIED	264908 264600, 264602, 264604, 264605, 264762, 264769, 264565
1114	14988014 (2227, 2228)	Novel Protein sim. GBank gi 1083428 pir S54876 - NAD(P)+ transhydrogenase (B-specific) (EC 1.6.1.1) precursor - mouse		dehydrogenase	264636
1115	11765583 (2229, 2230)				
1116	79841152 (2231, 2232)			UNCLASSIFIED	264686 264908

1117	85305465 (2233, 2234)	Novel Protein sim. GBank gi 3255955 emb CAA94089  - (270200) U5 snRNP-specific 200kD protein [Homo sapiens]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	18108392, 264488, 263394, 264489, 18108398, 56182575, 22278995, 22278996, 35696288, 56994075, 22278997, 22278998, 22278999, 264259, 26331822, 56182181, 29331824, 66714117, 29331825, 29331826, 60432289, 29331827, 29331828, 35696052, 33656970, 29146498, 264508, 264905, 264509, 264908, 264907, 264908, 66712502, 264909, 52644045, 56182435, 264510, 264511, 265008, 265007, 264512, 265008, 265009, 264910, 60170831, 264591, 264592, 264593, 60433356, 264594, 60433438, 264595, 55812038, 264596, 264758, 264759, 21906754, 33109954, 33657084, 265011, 87168559, 264600, 264601, 264602, 264603, 265017, 264604, 265018, 264605, 265019, 264760, 55811150, 264681, 264762, 18108351, 264448, 264682, 264764, 264683, 264288, 264389, 264684, 264685, 264766, 264767, 264686, 264687, 264768, 264769, 264688, 18108359, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 29148784, 35695917, 265020, 265021, 265022, 60170815, 264690, 264691, 33657023, 264692, 264693, 65274620, 33657109, 33657182, 27486261, 27486262, 27486264, 33657349, 27486265, 35695763, 264691
1118	79563326 (2235, 2236)		UNCLASSIFIED	264907
1119	79642463 (2237, 2238)		UNCLASSIFIED	264907
1120	79480463 (2239, 2240)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]	collagen	29331827, 265018, 265019, 264681, 265021, 60170615, 18108387
1121	79471716 (2241, 2242)	Novel Protein sim. GBank gi 1644450 (U67864) - MEX-3 [Caenorhabditis elegans]	Contains protein domain (PF00013) - KH domain	264683, 264632, 18108388
1122	79456246 (2243, 2244)		UNCLASSIFIED	264639, 264563
1123	79637119 (2245, 2246)	Novel Protein sim. GBank gi 98800 pir J17768 - 3-dehydroquinolate synthase (EC 4.6.1.3) - Mycobacterium tuberculosis	synthase	264693, 27486265
1124	79811596 (2247, 2248)		UNCLASSIFIED	264909
1125	79757861 (2249, 2250)		UNCLASSIFIED	264910
1126	79758914 (2251, 2252)	Novel Protein sim. GBank gi 138154 sp P03643 VGG_BPPHX - MAJOR SPIKE PROTEIN (G PROTEIN) (GPG)	eph	264905, 264909, 264910
1127	11800930 (2253, 2254)		UNCLASSIFIED	264682
1128	8364885 (2255, 2256)	Novel Protein sim. GBank gi 5002704 emb CAB44358.1  - (AJ242630) DNA polymerase I [Methylobacterium sp. DIM4]	Contains protein domain (PF00476) - DNA polymerase family A	264511

1129	80422480 (2257, 2258)	Novel Protein sim. GBank gi 5689485 dbj BAA03026.1  - (AB029997) KIAA1074 protein [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	265011, 264766
1130	79420151 (2259, 2260)	Novel Protein sim. GBank gi 4981328 gb AAD35881.1 AE00174 - (AE001747) bioY protein [Thermotoga maritima]		UNCLASSIFIED	264595
1131	80055391 (2261, 2262)	Novel Protein sim. GBank gi 4981328 gb AAD35881.1 AE00174 - (AE001747) bioY protein [Thermotoga maritima]		UNCLASSIFIED	35696286, 22278998, 29331828, 264603, 264605, 264559
1132	82062248 (2263, 2264)	Novel Protein sim. GBank gi 1841552 (U98336) - unknown [Homo sapiens]		UNCLASSIFIED	22278996, 264906, 265009, 264600, 264602, 264604, 264605, 264760, 32833986, 18108374
1133	17290437 (2265, 2266)			UNCLASSIFIED	265018
1134	80235376 (2267, 2268)			UNCLASSIFIED	264512, 264534
1135	80028383 (2269, 2270)	Novel Protein sim. GBank gi 4539171 emb CAB39700.1  - (AL049485) conserved hypothetical protein [Streptomyces coelicolor]		UNCLASSIFIED	264508, 264600, 264602, 264603, 18108376
1136	79842052 (2271, 2272)	Novel Protein sim. GBank gi 4982454 gb AAD36931.1 AE00182 - (AE001823) ATP-dependent protease LA, putative [Thermotoga maritima]		UNCLASSIFIED	264906, 264908
1137	90931557 (2273, 2274)	Novel Protein sim. GBank gi 4972746 gb AAD34768.1  - (AF132180) unknown [Drosophila melanogaster]	Contains protein domain (PF00515) - collagen TPR Domain		22278998, 22278999, 35696052, 264907, 265009, 60433356, 264596, 265010, 264448, 264682, 264767, 264689, 265020, 264692, 55811576, 35695855, 264631, 264632, 22278002
1138	79841163 (2275, 2276)	Novel Protein sim. GBank gi 731607 sp P38739 YHC8_YEAST - HYPOTHETICAL 63.8 KD PROTEIN IN GUT1-RIM1 INTERGENIC REGION PRECURSOR	strud		264908
1139	79633561 (2277, 2278)	Novel Protein sim. GBank gi 36550031 (AC005396) - putative proline-rich cell wall protein [Arabidopsis thaliana]		UNCLASSIFIED	264683
1140	39480358 (2279, 2280)			UNCLASSIFIED	264593
1141	79638019 (2281, 2282)			UNCLASSIFIED	265019, 264693
1142	19635848 (2283, 2284)			UNCLASSIFIED	264631
1143	87762158 (2285, 2286)	Novel Protein sim. GBank gi 3928000 emb CAA05880  - (AJ003125) procollagen I N-proteinase [Homo sapiens]	Contains protein domain (PF00090) - oxidase		56182575, 264908, 264600, 264632, 87188518
1144	80088988 (2287, 2288)		Thrombospondin type 1 domain	UNCLASSIFIED	264635, 264636, 264907, 264593, 264908, 264566, 264909
1145	14610262 (2289, 2290)			UNCLASSIFIED	264112
1146	82062092 (2291, 2292)			UNCLASSIFIED	264769, 264689, 35696286, 264760, 264905, 264486, 264559
1147	80071761 (2293, 2294)				264557
1148	80048433 (2295, 2296)	Novel Protein sim. GBank gi 2499003 sp P76422 THID_ECOLI - PHOSPHOMETHYL PYRIMIDINE KINASE (HMP-PHOSPHATE KINASE) (HMP-P KINASE)	kinase		264591
1149	11607438 (2297, 2298)	Novel Protein sim. GBank gi 2896734 emb CAA17213.1  - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]			264591

1150	81325074 (2289, 2300)	Novel Protein sim. GBank gjl2895095 (AF011337) - putative E1-E2 ATPase [Mus musculus]		ATPase associated	264488, 35696286, 264907, 264908, 264909, 264910, 264593, 264596, 264758, 264764, 264766, 264768, 264693, 264628, 60431850, 264564, 264566, 264567
1151	80070874 (2301, 2302)	Novel Protein sim. GBank gjl4324655 [gb]AAD16978] - (AF108191) DNA polymerase III alpha subunit [Streptomyces coelicolor]		polymerase	264595
1152	80235547 (2303, 2304)	Novel Protein sim. GBank gjl3874275 [emb]CAB07311.1] - (Z52825) predicted using GeneFinder: Similarity to Yeast low affinity glucose transporter HXT4 (PS:32467); cDNA EST EMBL: C12555 comes from this gene; cDNA EST yk404c10.3 comes from this gene; cDNA EST yk404c10.5 comes from this...		glycoprotein	264488, 22278998, 264905, 264629, 264486
1153	80027783 (2305, 2306)	Novel Protein sim. GBank gjl4240315 [dbj]BAA74936.1] - (AB020720) KIAA0913 protein [Homo sapiens]		UNCLASSIFIED	264910, 264555, 264557
1154	83002995 (2307, 2308)	Novel Protein sim. GBank gjl586655 [sp]P37617 [ATZN_ECOLI - ZINC-TRANSPORTING ATPASE (ZNF11)-TRANSLATING P. TYPE ATPASE]		UNCLASSIFIED	265008
1155	79411098 (2309, 2310)	Novel Protein sim. GBank gjl586655 [sp]P37617 [ATZN_ECOLI - ZINC-TRANSPORTING ATPASE (ZNF11)-TRANSLATING P. TYPE ATPASE]		UNCLASSIFIED	264690, 264636
1156	57147843 (2311, 2312)	Novel Protein sim. GBank gjl586655 [sp]P37617 [ATZN_ECOLI - ZINC-TRANSPORTING ATPASE (ZNF11)-TRANSLATING P. TYPE ATPASE]	Contains protein domain (PF00122) - E1-E2 ATPase	transport	264603
1157	95287711 (2313, 2314)	Novel Protein sim. GBank gjl418480 [sp]P32139 [YHR_ECOLI - HYPOTHETICAL 34.0 KD PROTEIN IN GLNA-RBN INTERGENIC REGION]		UNCLASSIFIED	264906, 264907, 264758, 264766, 264769, 264689, 264638, 264566
1158	82454917 (2315, 2316)	Novel Protein sim. GBank gjl249648 [sp]Q50724 [Y09S_MYCTU - HYPOTHETICAL 87.3 KD PROTEIN CY78.27C]			264906, 264762, 264687, 264769, 264689, 18108374, 35695855
1159	79186451 (2317, 2318)	Novel Protein sim. GBank gjl1136406 [dbj]BAA11490] - (D79995) similar to pig tubulin-tyrosine ligase. [Homo sapiens]			264687
1160	91228983 (2319, 2320)	Novel Protein sim. GBank gjl1136406 [dbj]BAA11490] - (D79995) similar to pig tubulin-tyrosine ligase. [Homo sapiens]			29331822, 29331824, 66714117, 29331826, 60433356, 265018, 265019, 83373044, 18108385, 22279000, 22279002, 264563
1161	7417143 (2321, 2322)	Novel Protein sim. GBank gjl2443342 [dbj]BAA22380] - (D88764) alpha 2 type I collagen [Rana catesbeiana]		UNCLASSIFIED	264369
1162	79635357 (2323, 2324)	Novel Protein sim. GBank gjl4503375 [ref]NP_001376.1 [pDPYS - dihydropyrimidinase]		UNCLASSIFIED	264693
1163	79563186 (2325, 2326)	Novel Protein sim. GBank gjl5052554 [gb]AAD38607.1 [AF14563] - (AF145632) BcDNA.GH06032 [Drosophila melanogaster]		transport	29331827, 264906
1164	79650829 (2327, 2328)	Novel Protein sim. GBank gjl5052554 [gb]AAD38607.1 [AF14563] - (AF145632) BcDNA.GH06032 [Drosophila melanogaster]			18108398, 29331827, 29331828, 29146488, 29146499, 18108354, 21906768, 29148627, 21906769, 264693, 18108382, 18108385
1165	80491888 (2329, 2330)	Novel Protein sim. GBank gjl4589476 [dbj]BAA76766.1] - (AB023139) KIAA0922 protein [Homo sapiens]		UNCLASSIFIED	264602, 264605, 264769, 18108370, 18108374, 264565
1166	88096456 (2331, 2332)	Novel Protein sim. GBank gjl4589476 [dbj]BAA76766.1] - (AB023139) KIAA0922 protein [Homo sapiens]			264488, 35696286, 22278999, 264259, 66714117, 60432289, 35696052, 264905, 56182435, 265006, 60433438, 264759, 21906754, 33109954, 265017, 265019, 244048, 264288, 264766, 264685, 35696423, 35695855, 264558, 18108385, 60432113

1167	78963862 (2333, 2334)	Novel Protein sim. GBank gij2580433[dbj BAA23138] - (D76414) ppGpp hydrolase [Staphylococcus aureus]			kinase	264488
1168	88094678 (2335, 2336)				UNCLASSIFIED	264259, 29331827, 56182435, 60433438, 265019, 33657023, 35695855, 264568
1169	11805403 (2337, 2338)				UNCLASSIFIED	264681
1170	21632244 (2339, 2340)				UNCLASSIFIED	264602
1171	20434582 (2341, 2342)	Novel Protein sim. GBank gij2772914 (AF029249) - precollagen D [Mytilus edulis]			UNCLASSIFIED	264556
1172	78610113 (2343, 2344)	Novel Protein sim. GBank gij4757846[ref NP_004317.1 pBCL9 - B-cell CLL/lymphoma 9			UNCLASSIFIED	55810764, 35696052, 264634, 264488
1173	80235713 (2345, 2346)	Novel Protein sim. GBank gij2564053[dbj BAA22946] - (AB007832) Bm trachealless [Bombyx mori]				264508, 264906, 264907, 264909, 264591, 264632, 264638, 264639
1174	20293077 (2347, 2348)	Novel Protein sim. GBank gij2911027[emb CAA117520] - (AL021958) mmsA [Mycobacterium tuberculosis]			dehydrogenase	264600
1175	20711847 (2349, 2350)	Novel Protein sim. GBank gij118333[sp P23234 DCIP_ENTCL - INDOL-3-PYRUVATE DECARBOXYLASE (INDOLEPYRUVATE DECARBOXYLASE)			carboxylase	264601
1176	80252645 (2351, 2352)	Novel Protein sim. GBank gij1144520 (U34956) - phosphoribosylformylglycinamide synthase [Mycobacterium tuberculosis]			synthase	264508, 264905, 264593, 264602, 264605
1177	80064647 (2353, 2354)	Novel Protein sim. GBank gij119791[sp P28643 FABG_CUPLA - 3-OXOACYL-(ACYL-CARRIER PROTEIN) REDUCTASE PRECURSOR (3-KETOACYL-ACYL CARRIER PROTEIN REDUCTASE)			reductase	264605
1178	94128641 (2355, 2356)	Novel Protein sim. GBank gij5031697[ref NP_005594.1 pFIC1 - familial intrahepatic cholestasis 1, (progressive, Byler disease and benign recurrent)			Contains protein domain (PF00122) - ATPase associated E1-E2 ATPase	65274572, 18108398, 22278998, 22278999, 29331826, 264508, 264908, 264828, 33657402, 33109954, 264769, 21806765, 21906766, 21806768, 55811957, 33657023, 264629, 55811576, 35696423, 264636, 264556, 56182323, 60432113, 22279000, 22279002
1179	80055575 (2357, 2358)	Novel Protein sim. GBank gij2960090[emb CAA17988.1] - (AL022121) dppA [Mycobacterium tuberculosis]			transport	264603
1180	11784446 (2359, 2360)	Novel Protein sim. GBank gij2558614[emb CAA04787] - (A001493) dehydroquinase dehydratase [Streptomyces coelicolor]			synthase	264638
1181	17946362 (2361, 2362)					
1182	81494284 (2363, 2364)	Novel Protein sim. GBank gij5420387[emb CAB46678.1] - (A1243459) proteophosphoglycan [Leishmania major]			UNCLASSIFIED	265017
1183	79574044 (2365, 2366)	Novel Protein sim. GBank gij4091877 (AF061331) - alpha galactosidase precursor [Saccharopolyspora erythraea]				264689, 35696423, 264638, 18108385
1184	52559933 (2367, 2368)	Novel Protein sim. GBank gij2129478[pir S1939 - chitinase (EC 3.2.1.14) precursor - beet			UNCLASSIFIED	264602
1185	79491185 (2369, 2370)				glycoprotein	263967



1186	20224012 (2371, 2372)				UNCLASSIFIED	264559
1187	79248834 (2373, 2374)				UNCLASSIFIED	29331825, 265017, 18108351
1188	79831367 (2375, 2376)	Novel Protein sim. GBank gij2996039 (AF054526) - hypothetical protein [Synechococcus PCC7002]			UNCLASSIFIED	264905, 264906
1189	79609367 (2377, 2378)					264692
1190	78930589 (2379, 2380)				UNCLASSIFIED	265018
1191	80310105 (2381, 2382)				UNCLASSIFIED	264600, 264605, 264764, 35695855, 264638, 264486
1192	13521641 (2383, 2384)		Contains protein domain (PF01581) - FMR family related peptide family			264636
1193	11103584 (2385, 2386)				UNCLASSIFIED	263978
1194	78833947 (2387, 2388)	Novel Protein sim. GBank gij854085 (emb)CAA583371 - (X83413) U88 [Human herpesvirus 6]			UNCLASSIFIED	265007, 265008
1195	20445442 (2389, 2390)	Novel Protein sim. GBank gij1790277 (AE000459) - putative oxidoreductase [Escherichia coli]			UNCLASSIFIED	264605
1196	13000688 (2391, 2392)					264689
1197	11392317 (2393, 2394)	Novel Protein sim. GBank gij2497360 (sp)Q50715 (IMDH_MYCTU - INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMP DEHYDROGENASE) (IMPDH) (IMPD)	Contains protein domain (PF00571) - CBS domain		dehydrogenase	264594
1198	95290101 (2395, 2396)					264603
1199	81882011 (2397, 2398)	Novel Protein sim. GBank gij1709525 (sp)P54673 (P3K1_DICDI - PHOSPHATIDYLINOSITOL 3-KINASE 1 (PI3-KINASE) (PTDINS-3-KINASE) (PI3K)				264259, 264757, 33109954, 21906768
1200	9848880 (2399, 2400)				UNCLASSIFIED	264910
1201	80503751 (2401, 2402)	Novel Protein sim. GBank gij2498877 (sp)P70645 (BLMH_RAT - BLEOMYCIN HYDROLASE (BLM HYDROLASE) (BMH)			cathepsin	264766, 264769
1202	80082633 (2403, 2404)	Novel Protein sim. GBank gij606342 (U18997) - ORF_o622; reading frame open far upstream of start; possible frameshift, linking to previous ORF [Escherichia coli]			ribosomal prot	264600, 264558
1203	82125373 (2405, 2406)				UNCLASSIFIED	264768, 264769, 35695917, 264910, 264760, 264908, 264907, 264629, 264908, 264909, 264766
1204	80503916 (2407, 2408)	Novel Protein sim. GBank gij2500728 (sp)Q59912 (SECY_STRGB - PREPROTEIN TRANSLOCASE SECY SUBUNIT				264905, 264769, 264636
1205	80053961 (2409, 2410)				UNCLASSIFIED	264566
1206	80241985 (2411, 2412)				UNCLASSIFIED	264556, 264557, 264558
1207	79841192 (2413, 2414)					29331824, 264909, 265021, 18108370
1208	87755217 (2415, 2416)	Novel Protein sim. GBank gij2645560 (AF027954) - Bcl-2-related ovarian killer protein [Rattus norvegicus]	Contains protein domain (PF00452) - Apoptosis regulator proteins, Bcl-2 family		apoptosis	29331824, 29331825, 29331827, 265007, 264764, 264683, 264769, 264688, 264689

1209	79185742 (2417, 2418)	Novel Protein sim. GBank gil1175033 sp P44398 XYLA_HAEIN - XYLOSE ISOMERASE	Contains protein domain (PF00259) - Xylose isomerase	isomerase	264687, 264688
1210	56426884 (2419, 2420)	Novel Protein sim. GBank hypothetical protein o246 - Escherichia coli		UNCLASSIFIED	264907, 264693
1211	94665655 (2421, 2422)	Novel Protein sim. GBank hypothetical protein o246 - Escherichia coli		transferase	264591, 264592, 264595
1212	79167929 (2423, 2424)	Novel Protein sim. GBank (293765) predicted using GeneFinder; similar to RNA recognition motif. (aka RRM, RBD, or RNP domain); cDNA EST EMBL: T01682 comes from this gene; cDNA EST EMBL: M75823 comes from this gene; cDNA EST EMBL: D27559 comes from this ge...	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf		264689, 263957
1213	79859633 (2425, 2426)	Novel Protein sim. GBank gene [Bordetella pertussis]		kinase	264909
1214	10144306 (2427, 2428)	Novel Protein sim. GBank gil5726285 gb AAD48396.1 AF12616 - (AF126162) HERV-H LTR associating protein 2 [Homo sapiens]		UNCLASSIFIED	264908
1215	80050108 (2429, 2430)	Novel Protein sim. GBank (Z98268) recN [Mycobacterium tuberculosis]		UNCLASSIFIED	265009, 264601, 264602, 264603, 33657109
1216	20438324 (2431, 2432)	Novel Protein sim. GBank gil417329 sp P33038 MURA_ENTCL - UDP-N- ACETYLGLUCOSAMINE 1- CARBOXYVINYL TRANSFERASE (ENOYL-PYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE ENOYL-PYRUVYL TRANSFERASE) (EPT)		transferase	264604
1217	95011344 (2433, 2434)			UNCLASSIFIED	264905, 264907, 264908, 264591, 264766, 264691, 264693, 264629, 264630, 264636, 264564
1218	11093680 (2435, 2436)	Novel Protein sim. GBank (D50453) homologue of succinate semialdehyde dehydrogenase GabD of E. coli [Bacillus subtilis]		dehydrogenase	264601
1219	91216252 (2437, 2438)				56181686, 29331822, 60432289, 264601, 264692, 264629
1220	91241524 (2439, 2440)	Novel Protein sim. GBank gil4240315 dbj BAA74936.1  - (AB020720) KIAA0913 protein [Homo sapiens]		oncogene	52844507, 264905, 264909, 265008, 265019, 265020, 52644150, 33657023, 264693, 33657182, 35695763, 264634, 22279000, 22279002, 264482
1221	83045055 (2441, 2442)	Novel Protein sim. GBank nucleoporin p62 homolog - rat (fragment)		UNCLASSIFIED	264768, 265020, 264906
1222	20711865 (2443, 2444)	Novel Protein sim. GBank gil730805 sp P39663 SPHR_SYN7 - ALKALINE PHOSPHATASE SYNTHESIS TRANSCRIPTIONAL REGULATORY PROTEIN SPHR	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	264601
1223	11615647 (2445, 2446)				264593
1224	80432645 (2447, 2448)	Novel Protein sim. GBank gil1172627 sp P46546 PROB_CORGL - GLUTAMATE 5- KINASE (GAMMA-GLUTAMYL KINASE) (GK)	Contains protein domain (PF01472) - PUA domain	kinase	264593, 264600, 264601, 264603, 264605, 264768, 18108376, 264635, 18108387

1225	80434427 (2449, 2450)	Novel Protein sim. GBank gij2105050[embjCAB08836] - (Z95436) hypothetical protein Rv3844c [Mycobacterium tuberculosis]			264768
1226	80237518 (2451, 2452)	Novel Protein sim. GBank gij1706768[spP98133]FBN1_BOVIN - FIBRILLIN 1 PRECURSOR (MP340)	polymerase		264905, 264512, 264689
1227	79422138 (2453, 2454)	Novel Protein sim. GBank gij1706768[spP98133]FBN1_BOVIN - FIBRILLIN 1 PRECURSOR (MP340)	UNCLASSIFIED		264908, 264637, 264639
1228	79209027 (2455, 2456)	Novel Protein sim. GBank gij1653901[dbjBAA18811] - (D90917) acriflavine resistance protein [Synechocystis sp.]			264605, 264634
1229	94329135 (2457, 2458)	Novel Protein sim. GBank gij116230[spP28598]CH60_BACSU - 60 KD CHAPERONIN (PROTEIN CPN60) [GROEL PROTEIN]	UNCLASSIFIED		87168474, 265011, 87168559, 264681, 264689, 264693, 65274620, 18108374
1230	80049357 (2459, 2460)	Novel Protein sim. GBank gij116230[spP28598]CH60_BACSU - 60 KD CHAPERONIN (PROTEIN CPN60) [GROEL PROTEIN]	UNCLASSIFIED	eph	264908, 264605, 18108388
1231	79843141 (2461, 2462)	Novel Protein sim. GBank gij1215733 (U48718) - OphC [Agrobacterium tumefaciens]	UNCLASSIFIED	transport	264908
1232	79853104 (2463, 2464)	Novel Protein sim. GBank gij116298[spP20730]CHHC_BOMMO - CHORION CLASS HIGH-CYSTEINE HCB PROTEIN 13 PRECURSOR (HC-B.13)	UNCLASSIFIED		264909
1233	80255179 (2465, 2466)	Novel Protein sim. GBank gij116298[spP20730]CHHC_BOMMO - CHORION CLASS HIGH-CYSTEINE HCB PROTEIN 13 PRECURSOR (HC-B.13)	UNCLASSIFIED		265017, 264564
1234	79242158 (2467, 2468)	Novel Protein sim. GBank gij129671[spP40280]H2A_MAIZE - HISTONE H2A		histone	265008, 265010, 18108381
1235	79914423 (2469, 2470)	Novel Protein sim. GBank gij3875133[embjCAA94750] - (Z70750) similar to actin binding domain; cDNA EST EMBL:U00093 comes from this gene; cDNA EST EMBL:D34443 comes from this gene; cDNA EST EMBL:D37508 comes from this gene; cDNA EST EMBL:D64247 comes from this gene; cDNA EST EMBL:...	UNCLASSIFIED		264634, 264762
1236	81927147 (2471, 2472)	Novel Protein sim. GBank gij3875133[embjCAA94750] - (Z70750) similar to actin binding domain; cDNA EST EMBL:U00093 comes from this gene; cDNA EST EMBL:D34443 comes from this gene; cDNA EST EMBL:D37508 comes from this gene; cDNA EST EMBL:D64247 comes from this gene; cDNA EST EMBL:...			265018, 55811150, 264565, 264757
1237	83371782 (2473, 2474)	Novel Protein sim. GBank gij3875133[embjCAA94750] - (Z70750) similar to actin binding domain; cDNA EST EMBL:U00093 comes from this gene; cDNA EST EMBL:D34443 comes from this gene; cDNA EST EMBL:D37508 comes from this gene; cDNA EST EMBL:D64247 comes from this gene; cDNA EST EMBL:...			264758, 264601, 264766, 264687, 18108372, 264555, 264559
1238	87411577 (2475, 2476)	Novel Protein sim. GBank gij3885470 (AF061443) - G protein-coupled receptor LGR4 [Rattus norvegicus]		glycoprotein	264259, 29331822, 29331824, 35696052, 264508, 264906, 52644045, 52646317, 264288, 264769, 264693, 264632, 264634, 264558, 87168518, 264563
1239	82197449 (2477, 2478)	Novel Protein sim. GBank gij4007990[gbjAAC95339] - (AF084363) DOK protein [Mus musculus]		oncogene	264509, 264511, 264759, 264760, 264764, 264557
1240	80497259 (2479, 2480)	Novel Protein sim. GBank gij1176192[spP45420]YHCD_ECOLI - HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN IN GLTF-NANT INTERGENIC REGION PRECURSOR			264769
1241	80020711 (2481, 2482)	Novel Protein sim. GBank gij121383[spP19904]GLNA_VIBAL - GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)	UNCLASSIFIED		264601, 264604, 264638
1242	79775890 (2483, 2484)	Novel Protein sim. GBank gij121383[spP19904]GLNA_VIBAL - GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)			264908, 264907, 264908, 264634

1243	79779458 (2485, 2486)	Novel Protein sim. GBank gij335567[emb]CAA19871] - (AL031124) branched-chain amino acid aminotransferase [Streptomyces coelicolor]		UNCLASSIFIED	18108374, 35695917, 35695955, 265009, 264508, 264909
1244	10284821 (2487, 2488)	Novel Protein sim. GBank gij2970846 (AF051945) - Xln [Mus musculus]		UNCLASSIFIED	264691
1245	80437103 (2489, 2490)	Novel Protein sim. GBank gij4586338[dbj]BAA76357.1] - (AB016787) cytochrome o ubiquinol oxidase B [Pseudomonas putida]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	264768
1246	80059321 (2491, 2492)	Novel Protein sim. GBank gij3581849[emb]CAA20805] - (AL031541) putative phenylalanyl-tRNA synthetase beta chain [Streptomyces coelicolor]		UNCLASSIFIED	264604, 264636, 264557, 264564
1247	80064831 (2493, 2494)	Novel Protein sim. GBank gij2621684 (AE000842) - adhesion protein [Methanobacterium thermoautotrophicum]			264758, 264605, 264639
1248	88070353 (2495, 2496)	Novel Protein sim. GBank gij1352403[sp]P09467F16P_HUMAN - FRUCTOSE-1,6-BISPHOSPHATASE (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPASE)	Contains protein domain (PF00316) - Fructose-1,6-bisphosphatase	UNCLASSIFIED	18108392, 264259, 29331826, 264106, 264508, 264907, 264828, 265009, 60433356, 264757, 264758, 21908754, 265010, 265011, 265018, 265019, 264760, 18108351, 18108354, 265021, 18108376, 18108377, 264630, 18108385
1249	80056657 (2497, 2498)	Novel Protein sim. GBank gij2791407[emb]CAA16001] - (AL021184) hypothetical protein Rv1473 [Mycobacterium tuberculosis]		transport	264908, 265010, 264600, 264603, 264691, 18108376
1250	12694385 (2499, 2500)	Novel Protein sim. GBank gij112785[sp]P051003MG1_ECOLI - DNA-3-METHYLADENINE GLYCOSIDASE I (3-METHYLADENINE-DNA GLYCOSYLASE I, CONSTITUTIVE) (TAG I)		UNCLASSIFIED	264689
1251	79850448 (2501, 2502)				264809
1252	79458087 (2503, 2504)			UNCLASSIFIED	264683, 263976
1253	80050121 (2505, 2506)	Novel Protein sim. GBank gij5670176[gb]AAD46616.1JAF16131 - (AF161317) NRAMP manganese transport protein MntA [Salmonella typhimurium]		glycoprotein	264600, 264603, 18108376
1254	87716767 (2507, 2508)	Novel Protein sim. GBank gij103160[pir]S22126 - finger protein unkempt - fruit fly (Drosophila melanogaster)		UNCLASSIFIED	35696286, 264910, 264764, 264688, 21906767, 55811957, 264692, 264556, 264639
1255	79168728 (2509, 2510)				264636
1256	87889508 (2511, 2512)	Novel Protein sim. GBank gij2995353[emb]CAA04608.1] - (AJ001206) pep2 [Streptomyces coelicolor]		UNCLASSIFIED	50432289, 264600, 264605, 264764, 264687, 264769, 264689, 27486265, 18108374, 18108376
1257	80201435 (2513, 2514)	Novel Protein sim. GBank gij3193306 (AF069300) - contains similarity to Arabidopsis membrane-associated salt inducible-like protein (GB:AL021637) [Arabidopsis thaliana]		UNCLASSIFIED	264094, 265019
1258	20708150 (2515, 2516)			UNCLASSIFIED	264602, 263978
1259	80186012 (2517, 2518)			UNCLASSIFIED	264906, 264448, 264908
1260	80084806 (2519, 2520)			UNCLASSIFIED	264634, 264639

1261	87412802 (2521, 2522)	Novel Protein sim. GBank gi 5689511 dbj BAAB3039.1  - (AB029010) KIAA1087 protein [Homo sapiens]	Contains protein domain (PF01699) - Sodium/calcium exchanger protein	cadherin	29331824, 264906, 264909, 264768, 264769, 264689, 264693, 264639, 18108384, 264563, 264634
1262	13504589 (2523, 2524)	Novel Protein sim. GBank gi 95100 pir S21334 - hypothetical protein 4 - Agrobacterium tumefaciens		UNCLASSIFIED	
1263	20710997 (2525, 2526)				264602
1264	80083396 (2527, 2528)	Novel Protein sim. GBank gi 3550958 (AF004840) - CDO [Rattus norvegicus]		struct	264634
1265	80253578 (2529, 2530)				264563
1266	79914604 (2531, 2532)			UNCLASSIFIED	264768, 264636, 264638, 264567
1267	80558918 (2533, 2534)	Novel Protein sim. GBank gi 1085002 pir S55056 - mitochondrial carrier protein DIF-1 homolog - Caenorhabditis elegans	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED - transport	264259, 21906754, 264369
1268	88178473 (2535, 2536)	Novel Protein sim. GBank gi 4886445 emb CAB43370.1  - (AL050269) hypothetical protein [Homo sapiens]	Contains protein domain (PF00583) - Acetyltransferase (GNAT) family	UNCLASSIFIED	18108398, 22278995, 56994075, 60424269, 29331827, 264108, 264512, 265007, 265008, 265009, 264595, 33109954, 33657084, 87168559, 264600, 265018, 265019, 264369, 264688, 21906767, 265020, 52644150, 264691, 33657023, 33657349, 18108374, 264556, 18108385, 60432113, 22279002, 264486
1269	78821946 (2537, 2538)	Novel Protein sim. GBank gi 3334791 emb CAA1939  - (AL031107) hypothetical protein SC5A7.10c [Streptomyces coelicolor]		UNCLASSIFIED	264508, 264905, 264906, 264687, 264693
1270	80031420 (2539, 2540)	Novel Protein sim. GBank gi 2851634 sp Q50591 Y001_MYCTU - HYPOTHETICAL 50.0 KD PROTEIN CY1A11.01	Contains protein domain (PF01574) - IMP dehydrogenase / GMP reductase N terminus	dehydrogenase	265010, 264601
1271	78840498 (2541, 2542)			ATPase associated	35696052, 264908
1272	79462878 (2543, 2544)				264686, 264689
1273	80220315 (2545, 2546)	Novel Protein sim. GBank gi 1655665 emb CAB03731  - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]		UNCLASSIFIED	264509, 264639
1274	95010802 (2547, 2548)			UNCLASSIFIED	
1275	20730763 (2549, 2550)	Novel Protein sim. GBank gi 123726 sp P10413 HTPG_ECOLI - HEAT SHOCK PROTEIN HTPG (HIGH TEMPERATURE PROTEIN G) (HEAT SHOCK PROTEIN C82.5)	Contains protein domain (PF00183) - Hsp90 protein	eph	264805, 264908, 264909, 264769, 264802
1276	21148644 (2551, 2552)	Novel Protein sim. GBank gi 2129478 pir S51939 - chitinase (EC 3.2.1.14) precursor - beet		UNCLASSIFIED	264369
1277	20438195 (2553, 2554)				
1278	11088365 (2555, 2556)	Novel Protein sim. GBank gi 1175473 sp P4455 YAAJ_HAEIN - HYPOTHETICAL PROTEIN H10183		UNCLASSIFIED	264556, 264603
1278	21658756 (2557, 2558)	Novel Protein sim. GBank gi 1929513 (U64316) - ATP synthase subunit beta [Mordella thermocetica]		UNCLASSIFIED	
1280	79310959 (2559, 2560)	Novel Protein sim. GBank gi 4938504 emb CAB43862.1  - (AL078465) putative protein [Arabidopsis thaliana]		synthase	264605
				struct	263876

1281	94323988 (2561, 2562)	Novel Protein sim. GBank gij1136501 (U39546) - surface protein MCA-32 [Rattus norvegicus]	Contains protein domain (PF00047) - Immunoglobulin domain	UNCLASSIFIED	29331825, 29331828, 264766, 83373044
1282	87537695 (2563, 2564)	Novel Protein sim. GBank gij3328180 (AF074266) - proto-oncogene AF4 [Mus musculus]		UNCLASSIFIED	265008
1283	20466305 (2565, 2566)	Novel Protein sim. GBank gij3261721(emb)(CAB07057) - (Z92770) hypothetical protein Rv0153c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605
1284	20636325 (2567, 2568)	Novel Protein sim. GBank gij3929022 (AF057696) - LspB [Haemophilus ducreyi]			264604
1285	80427330 (2569, 2570)	Novel Protein sim. GBank gij4171541spjP33126jHS82_ORYSA - HEAT SHOCK PROTEIN 82	Contains protein domain (PF00183) - Hsp90 protein	eph	264766, 264689, 263967
1286	20465254 (2571, 2572)	Novel Protein sim. GBank gij2078004(emb)(CAB08451) - (Z95207) gorA [Mycobacterium tuberculosis]		reductase	264605, 264639
1287	80417530 (2573, 2574)			UNCLASSIFIED	265011, 264602, 264766, 264687, 264769, 264689, 18108370, 264636, 18108385, 264563
1288	95338101 (2575, 2576)	Novel Protein sim. GBank gij5353510jgbIAA.D42161.1(AF08891) - (AF088916) emilin precursor [Homo sapiens]	Contains protein domain (PF00386) - C1q domain	collagen	35696052, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265007, 264512, 264910, 265009, 33657402, 264595, 264758, 265011, 265019, 264760, 18108351, 264681, 264764, 264288, 264685, 264766, 264687, 264768, 264769, 265020, 265021, 264534, 264692, 18108370, 264628, 18108374, 35696423, 264555, 264556, 264557, 264558, 18108385, 264564, 264566, 264567, 264486, 18108391
1289	11813647 (2577, 2578)	Novel Protein sim. GBank gij1169985jP46023GPCR_LYMST - G-PROTEIN COUPLED RECEPTOR GRL101 PRECURSOR	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	UNCLASSIFIED	264837
1290	19526027 (2578, 2580)	Novel Protein sim. GBank gij2072674(emb)(CAB08305) - (Z95120) rhlE [Mycobacterium tuberculosis]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	Im7	264563
1291	80470266 (2581, 2582)	Novel Protein sim. GBank gij1835755 (U86338) - zinc finger protein Png-1 [Mus musculus]	Contains protein domain (PF01530) - Zinc finger, C2HC type	helicase	265007, 265008, 264769
1292	94723316 (2583, 2584)	Novel Protein sim. GBank gij2129173jpirjF64453 - oxalacetate decarboxylase (EC 4.1.1.3) alpha subunit - Methanococcus jannaschii		transcriptfactor	264092, 264259, 29331822, 29331824, 264508, 264908, 264909, 264512, 265008, 265009, 264591, 265019, 264369, 264288, 264686, 264768, 264693, 18108374, 264632, 56182323, 264639, 83373044, 22279002, 264482, 264563
1293	80667536 (2585, 2586)	Novel Protein sim. GBank gij2129173jpirjF64453 - oxalacetate decarboxylase (EC 4.1.1.3) alpha subunit - Methanococcus jannaschii		UNCLASSIFIED	265008, 55812038, 264369, 264556
1294	82125908 (2587, 2588)	Novel Protein sim. GBank gij5441779(emb)(CAB46803.1) - (AL096811) putative alcohol dehydrogenase (zinc-binding) [Streptomyces coelicolor A3(2)]		biotindep	264602, 264605, 264760, 18108351, 264689, 33657023, 264559
1295	11686851 (2589, 2590)			dehydrogenase	264689

1296	11687904 (2591, 2592)	Novel Protein sim. GBank			UNCLASSIFIED	264591, 264639
1297	78639300 (2593, 2594)	gl14982191gb AAD36866.1 AE00180 - (AE001805) DNA-directed DNA polymerase I [Thermotoga maritima]	Contains protein domain (PF01367) - 5'-3' exonuclease		polymerase	264693
1288	94239506 (2595, 2596)	Novel Protein sim. GBank gl11943770 (U97191) - F53F10.1 gene product [Caenorhabditis elegans]			struct	18108348, 265017
1289	80255378 (2597, 2598)					
1300	80064967 (2599, 2600)	Novel Protein sim. GBank gl13445181 (AC005498) - R31665.2 [Homo sapiens]	Contains protein domain (PF01352) - KRAB box		transcriptfactor	264488, 264906, 264909, 22279002, 264566, 264605
1301	17839614 (2601, 2602)	Novel Protein sim. GBank gl14062973dbj BAA36204.1 - (AB017138) alpha subunit of malonate decarboxylase [Pseudomonas putida]			UNCLASSIFIED	264906
1302	95416188 (2603, 2604)					85658542, 265020
1303	9684121 (2605, 2606)					264908
1304	79377196 (2607, 2608)				UNCLASSIFIED	264508
1305	19905899 (2609, 2610)					264566
1306	13069230 (2611, 2612)	Novel Protein sim. GBank gl13242273jemb CAB070171 - (Z82669) hypothetical protein Rv0236c [Mycobacterium tuberculosis]			UNCLASSIFIED	264636
1307	82201029 (2613, 2614)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketoacyl acyl carrier protein reductase [Bacillus subtilis]	Contains protein domain (PF00516) - Envelope glycoprotein GP120		UNCLASSIFIED	264907, 264592, 264764
1308	21426814 (2615, 2616)	Novel Protein sim. GBank gl195819ipir S16298 - ferric enterobactin transport protein fepC - Escherichia coli			reductase	264555
1309	78263011 (2617, 2618)	Novel Protein sim. GBank gl15459220jemb CAB48993.1 - (AL096837) putative iron-sulfur protein [Streptomyces coelicolor A3(2)]			transport	264906, 18108354
1310	20466319 (2619, 2620)	Novel Protein sim. GBank gl14455118jgb AAD210841 - (AF125158) zinc finger DNA binding protein 99 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type		UNCLASSIFIED	264605
1311	87613142 (2621, 2622)					35696286, 29331827, 264908, 265008, 264764, 264766, 264686, 21908767, 21908769, 35695917, 264691, 264693, 22278995, 22278998, 22278999, 264905, 264908, 265011, 265017, 285019, 264687, 21906768, 265020, 265021, 33657023, 22279002, 264564
1312	88061720 (2623, 2624)	Novel Protein sim. GBank gl14455118jgb AAD210841 - (AF125158) zinc finger DNA binding protein 99 [Homo sapiens]			dna_ma_bind	22278996, 22278999, 264259, 20281099, 29146498, 264508, 264908, 66712502, 60433356, 60433438, 265011, 265017, 284683, 264288, 21906765, 21806767, 29148627, 21906768, 35695917, 265021, 33657023, 33657109, 18108370, 18108377, 35695855, 60432113, 22279000, 264563, 18108390
1313	91225458 (2625, 2626)	Novel Protein sim. GBank gl14929733jgb AAD34127.1 AF15189 - (AF151890) CGI-132 protein [Homo sapiens]	Contains protein domain (PF00886) - Ribosomal protein S16		ribosomalprot	264693
1314	56926053 (2627, 2628)					264691
1315	84357192 (2629, 2630)	Novel Protein sim. GBank gl12589223 (AF026565) - ring finger protein [Mus musculus]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)		interleukinrecept	

1316	95361609 (2631, 2632)	Novel Protein sim. GBank gji5689407[dbj][BAA82987.1] - (AB028958) KIAA1035 protein [Homo sapiens]		kinase	56182575, 56181686, 20281171, 29331822, 29331824, 60424268, 29331825, 35696052, 52644045, 264591, 60432229, 265018, 265019, 55811150, 56181562, 21906765, 21906787, 21906768, 35695917, 60170815, 33657023, 65274620, 33657109, 35695763, 35695855, 18108387, 87168518, 60432113, 22279002, 264564
1317	88055167 (2633, 2634)	Novel Protein sim. GBank gji4836757[gb][AAD30541.1][AF134918] semaphorin subclass 4 member G [Mus musculus]		UNCLASSIFIED	264093, 264806, 264908, 264369, 264884
1318	95322893 (2635, 2636)	Novel Protein sim. GBank gji4680204[gb][AAD27567.1][AF114171] hypothetical protein [Sorghum bicolor]		UNCLASSIFIED	18108392, 18108348, 265011, 265017, 18108359, 18108362, 56182323, 18108385, 22279000
1319	94236546 (2637, 2638)				264908, 264909, 265006, 265008, 264592, 265019, 264766, 56181562, 18108368, 264628, 264629, 18108377, 264636
1320	86603587 (2639, 2640)	Novel Protein sim. GBank gji4240183[dbj][BAA74870.1] - (AB020654) KIAA0847 protein [Homo sapiens]		UNCLASSIFIED	35696286, 55812038, 265018, 21906768, 265020, 263978, 22279002
1321	86676351 (2641, 2642)	Novel Protein sim. GBank gji4886505[emb][CAB43377.1] - (AL050276) hypothetical protein [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	transcript factor	60432049, 29331828, 264907, 264908, 264909, 264910, 55812038, 264601, 264762, 264764, 264766, 264768, 264769, 264628, 18108374, 264634, 264635, 18108385
1322	87756272 (2643, 2644)	Novel Protein sim. GBank gji526259[emb][CAB45736.1] - (AL080143) hypothetical protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_rna_bind	29331828, 264908, 265020, 33657023, 264693, 264404
1323	94845931 (2645, 2646)	Novel Protein sim. GBank gji5459516[dbj][BAA82407.1] - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]		synthase	65274572, 56994075, 264259, 29331822, 29331827, 264104, 56182435, 87168474, 18108351, 264288, 21906766, 21906767, 35695917, 265020, 264693, 65274791, 56182323, 18108387
1324	87737614 (2647, 2648)	Novel Protein sim. GBank gji5031717[ref][NP_005704.1]pGPBP - goodpasture antigen-binding protein	Contains protein domain (PF01852) - START domain		22278996, 22278998, 29331828, 264905, 264907, 29331830, 264908, 264510, 265008, 264595, 264759, 21906754, 265018, 264288, 264768, 264769, 21906768, 265022, 18108376, 264631, 264632, 264634, 264636, 264638, 264563, 264564, 264565, 264566
1325	94847471 (2649, 2650)	Novel Protein sim. GBank gji3294501 (U64857) - similar to the DPTUKunitz family of inhibitors; most similar to tissue factor pathway inhibitor precursor [Caenorhabditis elegans]	Contains protein domain (PF00090) - Thrombospondin type 1 domain	protease	35696286, 264905, 264906, 264907, 264908, 264909, 264910, 264593, 33657402, 264758, 8558542, 264760, 264768, 264769, 264691, 35698423
1326	87316289 (2651, 2652)	Novel Protein sim. GBank gji1397275 (U61947) - C06G3.8 gene product [Caenorhabditis elegans]		UNCLASSIFIED	264259, 66712502, 264682, 264683, 264635



1327	95322897 (2653, 2654)	Novel Protein sim. GBank gi 728832 sp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII	Contains protein domain (PF00279) - Plant lipid transfer protein family	UNCLASSIFIED	18108398, 22278996, 22278997, 22278999, 264091, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264105, 264905, 56182435, 264112, 265008, 265009, 21906754, 265010, 265011, 265017, 265019, 264681, 264448, 264764, 264684, 264288, 264685, 264768, 264688, 21906767, 21906769, 29148629, 285020, 284690, 264691, 264692, 264693, 263967, 33657109, 33657182, 27486262, 33657349, 18108370, 18108374, 55810764, 35695855, 264634, 56182323, 83373044, 87168518, 60432113, 22279000, 22279002, 264553, 264567
1328	87753493 (2655, 2656)			UNCLASSIFIED	264488, 22278997, 29331826, 264595, 18108351, 264766, 22279002, 264482, 264567
1329	87755276 (2657, 2658)	Novel Protein sim. GBank gi 4678224 gb AAD26969.1 AC00713 - (AC007135) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	22278996, 29331827, 264684, 264692, 33657109
1330	87727737 (2659, 2660)	Novel Protein sim. GBank gi 437310 (L23504) - nodulin [Medicago truncatula]		UNCLASSIFIED	264259, 29331825, 264512, 265019, 265021, 264555, 264558, 56528486
1331	87376764 (2661, 2662)	Novel Protein sim. GBank gi 4589586 dbj BAA76815.1  - (AB023188) KIAA0971 protein [Homo sapiens]		UNCLASSIFIED	264259, 29331826, 29331827, 35696052, 29331828, 60170831, 264448, 264686, 21906765, 55811857, 265020, 33657023, 33657109, 263973, 55811576, 35696423, 35695855, 56182323
1332	94845937 (2663, 2664)	Novel Protein sim. GBank gi 5459516 dbj BAA82407.1  - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]		synthase	55274572, 22278996, 56994075, 22278999, 60432049, 264259, 29331822, 29331826, 60432289, 29331827, 35696052, 52644045, 56182435, 264510, 21906754, 87168559, 265018, 265019, 264448, 264288, 264369, 264686, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 52644150, 33657023, 56182323, 18108387, 60432113, 22279002
1333	88098476 (2665, 2666)	Novel Protein sim. GBank gi 5689527 dbj BAA83047.1  - (AB028018) KIAA1095 protein [Homo sapiens]		UNCLASSIFIED	60432289, 66712502, 264591, 60433356, 60433438, 55812038, 265010, 264639, 56528486
1334	87592388 (2667, 2668)	Novel Protein sim. GBank gi 2662536 (AF036685) - Similar to protein-tyrosine phosphatase [Caenorhabditis elegans]		phosphatase	264905
1335	87644798 (2669, 2670)	Novel Protein sim. GBank gi 4240285 dbj BAA74921.1  - (AB020705) KIAA0898 protein [Homo sapiens]	Contains protein domain (PF00843) - B-box zinc finger.	UNCLASSIFIED	22278998, 22278999, 29331827, 264509, 264511, 265007, 265008, 265009, 60433438, 21906754, 87168559, 265017, 284288, 21906765, 21906767, 21906768, 21906769, 265020, 33657109, 27486264, 18108374, 264556, 264638, 264557, 60170394, 264559, 18108385, 264563

1336	87787890 (2671, 2672)	Novel Protein sim. GBank gi 465445 sp P33485 VNUA_PRVKA - PROBABLE NUCLEAR ANTIGEN			UNCLASSIFIED	264508, 264905, 264512, 264764, 264683, 264635, 264637
1337	94312042 (2673, 2674)	Novel Protein sim. GBank gi 5689471 dbj BAA83019.1  - (AB028990) KIAA1067 protein [Homo sapiens]			UNCLASSIFIED	56182575, 56994075, 22278998, 22278999, 264092, 264259, 60432289, 29331826, 264906, 264908, 264909, 264112, 265008, 265009, 60433356, 55812038, 33657084, 265011, 265017, 265018, 265019, 264682, 264448, 264683, 264369, 264688, 264689, 21906786, 21906789, 265020, 264691, 27486261, 20281069, 18108379, 55811576, 35695855, 55182323, 60432113, 22279002, 264567
1338	80366114 (2675, 2676)				UNCLASSIFIED	29331822, 265010, 264288, 264689, 18108370, 35695855
1339	80249231 (2677, 2678)	Novel Protein sim. GBank gi 1176422 (U43194) - rhophilin [Mus musculus]			UNCLASSIFIED	35696052, 264909, 264688, 264556, 264558
1340	88316311 (2679, 2680)					264905, 264907, 87168559, 264764
1341	86101485 (2681, 2682)					264681, 264685, 264686, 264692
1342	80089017 (2683, 2684)	Novel Protein sim. GBank gi 5019564 emb CAB44507.1  - (AL035542) dJ994E9.5 (hs6M1-17 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein)) [Homo sapiens]	Contains protein domain (PF000001) - tm7 7 transmembrane receptor (rhodopsin family)			264629
1343	80082862 (2685, 2686)	Novel Protein sim. GBank gi 4557543 ref NP_001384.1 pECM2 - extracellular matrix protein 2	Contains protein domain (PF00560) - struct Leucine Rich Repeat			264810, 264686, 264534
1344	20562559 (2687, 2688)	Novel Protein sim. GBank gi 2144101 pir I55210 - tricarboxylate carrier - rat (fragment)			glycoprotein	263978
1345	91225546 (2689, 2690)	Novel Protein sim. GBank gi 3881052 emb CAA19523  - (AL023843) predicted using GeneFinder, similar to serine/threonine kinase; cDNA EST yk246a12.3 comes from this gene; cDNA EST yk358c10.5 comes from this gene; cDNA EST EMBL:M89047 comes from this gene; cDNA EST yk246a12.5 comes...	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain			22278998, 264907, 264681, 264685, 264689, 265020, 264683, 22279000, 22279002, 264566
1347	80417393 (2693, 2694)	Novel Protein sim. GBank gi 4504379 ref NP_003658.1 pHG38 - orphan G protein- coupled receptor HG38	Contains protein domain (PF00560) - Leucine Rich Repeat		glycoprotein	264908, 264908, 264909, 265006, 264910, 265011, 265017, 264764, 264766, 264767, 264769, 264631, 264634, 264638, 264567, 264486
1348	87352335 (2695, 2696)	Novel Protein sim. GBank gi 3399720 dbj BAA32100  - (AB010999) peptidylarginine deiminase type IV [Rattus norvegicus]			UNCLASSIFIED	264488, 264489, 264508, 264509, 264510, 264511, 264512, 264591, 264592, 264601, 264684, 264685, 264769, 264532, 264534, 264555, 264556, 264557, 264558, 22279002, 264486

1349	91225548 (2687, 2688)	Novel Protein sim. GBank gij2144101 pir  55210 - tricarboxylate carrier - rat (fragment)		UNCLASSIFIED	52646842, 35695286, 22278996, 22278998, 22278999, 264259, 29331822, 29331824, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265008, 264512, 264910, 60170831, 264591, 60433438, 264757, 21906754, 265017, 265018, 264605, 264760, 264762, 264288, 264768, 264689, 21906765, 21906766, 21906767, 21906768, 55811957, 35695917, 265020, 264534, 264691, 264692, 33657023, 264693, 33657349, 18108374, 18108378, 35696423, 60170394, 22279000, 22279002, 264563, 264564
1350	87083136 (2689, 2700)			UNCLASSIFIED	52646842, 264259, 29331825, 264908, 264511, 264604, 264288, 21906769, 265020, 33657182, 33657349, 18108374, 35695855, 264555, 264558, 18108385, 22279002, 264486
1351	87361327 (2701, 2702)	Novel Protein sim. GBank gij4887239 gb AAD32246.1  - (AF064564) BAW protein [Fugu rubripes]		UNCLASSIFIED	264906, 264907, 264638
1352	80076386 (2703, 2704)			UNCLASSIFIED	264693, 263981
1353	95345417 (2705, 2708)	Novel Protein sim. GBank gij2144101 pir  55210 - tricarboxylate carrier - rat (fragment)		UNCLASSIFIED	35696286, 60424269, 264905, 264509, 264906, 264907, 264908, 264909, 264511, 264512, 264910, 264758, 264598, 55811386, 265011, 264605, 55811150, 264762, 264784, 264766, 52644229, 56181562, 35695917, 265022, 33657023, 264693, 35695763, 60431528, 264629, 263978, 35696423, 35695855, 264630, 264634, 264635, 264636, 264637, 264638, 264639, 18108385, 264563, 264564, 264566
1354	95350845 (2707, 2708)	Novel Protein sim. GBank gij4889108 gb AAD27763.1 AF07703 - (AF077030) hypothetical 43.2 kDa protein [Homo sapiens]		UNCLASSIFIED	22278995, 22278999, 29331826, 264806, 265008, 33657402, 21906754, 265011, 87188559, 264684, 264369, 264769, 264689, 21906765, 21906768, 52644150, 33657023, 264692, 264693, 18108374, 83373044, 87188518, 22279000
1355	88260186 (2709, 2710)	Novel Protein sim. GBank gij1469199 dbj BAA08487  - (D50828) The KIAA0138 gene product is novel. [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	22278996, 22278997, 264259, 66714117, 264511, 21906754, 265010, 264769, 264689, 21906765, 21906768, 21906769, 264532, 27486282, 264629, 264638, 264556, 264638, 264639, 264482, 264484

1356	85313991 (2711, 2712)	Novel Protein sim. GBank gij113865 (U40342) - ninein [Mus musculus]		struct	18108397, 22278995, 22278998, 22278998, 264094, 29331828, 264905, 265008, 265007, 265008, 265010, 265017, 265018, 265019, 264764, 18108354, 264689, 21906765, 265022, 18108384, 35696423, 83373044, 18108387
1357	88260268 (2713, 2714)	Novel Protein sim. GBank gij897693[emb]CAA90330] - (Z50026) phosphatidylcholine transfer protein [Bos taurus]	Contains protein domain (PF01852) - START domain		264259, 29331822, 29331825, 264510, 87168559, 265018, 264448, 264288, 21908765, 21908768, 21908768, 265021, 264693, 18108376
1358	38718455 (2715, 2716)	Novel Protein sim. GBank gij556219 (L36831) - transcription regulator [Mus musculus]			264757
1359	87771643 (2717, 2718)			UNCLASSIFIED	264907, 264909, 264510, 264511, 264512, 18108351, 264764, 264534, 33657023, 18108374, 264634, 264635, 264638, 264639, 18108385, 264488, 264587
1360	87738272 (2719, 2720)	Novel Protein sim. GBank gij2589282[emb]CAA75612] - (Y15417) acetate-CoA ligase [Coprinus cinereus]		synthase	60432289, 264605
1361	87593527 (2721, 2722)	Novel Protein sim. GBank gij5889443[db]BAA83005.1] - (AB028976) KIAA1053 protein [Homo sapiens]	Contains protein domain (PF00536) - SAM domain (Sterile alpha motif)	UNCLASSIFIED	35696286, 22278997, 22278999, 264259, 29331826, 264508, 264509, 264905, 264907, 264908, 265007, 265009, 33109954, 21908754, 87168474, 265011, 264761, 264683, 264288, 264766, 264769, 264689, 21906768, 265020, 265021, 33657023, 55811576, 35696423, 264634, 60432113, 22279002, 264482, 264486
1362	95287961 (2723, 2724)	Novel Protein sim. GBank gij568941[db]BAA82989.1] - (AB028960) KIAA1037 protein [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	eph	56182575, 56181686, 60432049, 264259, 29331822, 56182181, 29331827, 35696052, 29331828, 264905, 264908, 264908, 264595, 55812038, 85658542, 55811150, 264681, 264288, 264369, 56181562, 60431528, 55810764, 35696423, 60431850, 284558
1363	85758476 (2725, 2726)	Novel Protein sim. GBank gij1130494 (U35776) - ADP-ribosylation factor 1-directed GTPase activating protein [Rattus norvegicus]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	UNCLASSIFIED	264488, 29331826, 264907, 264687, 264689, 264693
1364	88179488 (2727, 2728)				60432289, 60433356, 60433438, 87168559, 264603, 18108351, 21908768, 35696423, 60432113
1365	83003108 (2728, 2730)	Novel Protein sim. GBank gij4589562[db]BAA76803.1] - (AB023176) KIAA0959 protein [Homo sapiens]		oncogene	264766
1366	87003262 (2731, 2732)	Novel Protein sim. GBank gij1084944[db]S54495 - hypothetical protein YPR021c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	265007

1367	87721210 (2733, 2734)	Novel Protein sim. GBank gij4884088[emb](CAB43240.1) - (AL050018) hypothetical protein [Homo sapiens]	Contains protein domain (PF01342) - SAND domain	UNCLASSIFIED	264488, 52646842, 52646365, 22278995, 56994075, 35698286, 22278996, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 60432289, 35696052, 264805, 264907, 264908, 264909, 264510, 265006, 265007, 264512, 264910, 265009, 60170831, 33657402, 55812038, 21906754, 265011, 87168559, 265017, 265019, 18108351, 264448, 264682, 264683, 264288, 264369, 264886, 264767, 264688, 21906765, 21906766, 21906768, 21906769, 55811957, 265020, 265021, 265022, 264534, 60170815, 264690, 264691, 18108362, 33657023, 33657109, 33657349, 264628, 18108370, 18108374, 18108376, 55811576, 35696423, 35695855, 264635, 264555, 264637, 264556, 52844332, 60170394, 264558, 18108381, 18108385, 56526486, 22279000, 264563, 264587
1368	94320078 (2735, 2738)	Novel Protein sim. GBank gij464561[sp]P35289[RAB15_RAT RAS-RELATED PROTEIN RAB-15]	Contains protein domain (PF00071) - Ras family	oncogene	264259, 29331822, 29331826, 60432289, 29331827, 35696052, 264508, 264905, 264906, 264908, 264909, 264510, 265007, 264910, 60433438, 264756, 85636542, 87168559, 264600, 264601, 264760, 264764, 264765, 264768, 52644229, 264689, 35695917, 265020, 265021, 264631, 264632, 264634, 264637, 52644332, 264558, 264639, 83373044, 264563, 264586, 264486, 264567, 265008, 60432229, 60433356, 33657084, 21906764, 21906768, 264555, 264638, 264559, 264567
1369	86634033 (2737, 2738)	Novel Protein sim. GBank gij2062702 (U90550) - butyrophilin [Homo sapiens]		UNCLASSIFIED	
1370	95316910 (2739, 2740)	Novel Protein sim. GBank gij5031823[ref]NP_005823.1[KCNM - potassium large conductance calcium-activated channel, subfamily M, beta member 2]		potassium_channel	22278996, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264758, 265011, 265018, 264764, 264766, 264768, 21906767, 35695917, 18108362, 35696423, 264632, 264635, 264636, 264555, 264638, 264558, 264639, 18108385, 65274727, 264404, 264563, 264568, 264486
1371	95336512 (2741, 2742)	Novel Protein sim. GBank gij5032203[ref]NP_005714.1[TPSPA - tetraspan 5]	Contains protein domain (PF00335) - 4 transmembrane segments integral membrane proteins	glycoprotein	22278996, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 264905, 264509, 29331830, 66712502, 265008, 265009, 264758, 33657084, 85658542, 265010, 265018, 265019, 264762, 264448, 35695917, 33657109, 33657182, 33657349, 35695855, 264558, 22279002, 264563

1372	80248517 (2743, 2744)	Novel Protein sim. GBank gij840708[dbj BAA09334] - (D50685) trans-sialidase [Trypanosoma cruzi]		collagen	263978
1373	80489421 (2745, 2746)			UNCLASSIFIED	264769, 21906765, 21906767, 22278999, 264691, 264910, 55812038, 265010, 264681, 264684
1374	95087036 (2747, 2748)	Novel Protein sim. GBank gij111876[pir JC1241] - beta-interferon-induced protein - rat		interferon	264907, 264510, 265007, 264512, 265008, 60432229, 264689, 65274791, 264555, 264556, 264557, 83373044, 60432113
1375	94236942 (2749, 2750)	Novel Protein sim. GBank gij5649176[gb AAD03500.2] - (AF051155) G beta-like protein GBL [Rattus norvegicus]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	52644507, 52645156, 52646842, 52646365, 56182575, 56181686, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 52645080, 29331824, 29331825, 60432229, 29331826, 29331827, 29331828, 35696052, 33656970, 264905, 264509, 264906, 264907, 264908, 29331830, 264909, 265006, 264511, 265007, 285008, 265009, 33657402, 60433356, 52646317, 33109954, 33657084, 52644296, 87168474, 87168559, 264600, 265017, 265018, 285019, 55811150, 18108351, 264448, 264764, 264288, 264369, 264766, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 52644150, 33657023, 264693, 65274620, 52645129, 33657109, 27486261, 33657349, 27486265, 35695763, 18108376, 55810764, 35696423, 35695855, 264630, 264631, 264634, 264636, 264555, 264638, 18108385, 87168518, 60432113, 22279000, 22279002, 264563, 264564, 264566, 264567, 264768, 264769, 35695917, 22278997, 264691, 264259, 29331822, 264693, 35696052, 264508, 264509, 264905, 264906, 264628, 264908, 264629, 18108372, 264809, 264510, 264511, 264512, 265008, 264630, 264631, 264910, 264632, 264634, 264635, 264636, 264591, 264592, 264637, 264638, 264558, 264639, 33657402, 264595, 18108385, 56526486, 265010, 265011, 264600, 264563, 264762, 264564, 264565, 264764, 264486, 264766
1376	87399050 (2751, 2752)	Novel Protein sim. GBank gij138350[sp P28968 VGLX_HSVB - GLYCOPROTEIN X PRECURSOR		UNCLASSIFIED	28331824, 264591, 265019, 264686, 264768, 55811957, 264693, 22279002
1377	86864242 (2753, 2754)	Novel Protein sim. GBank gij1663648 (U75321) - chromaffin granule ATPase II homolog [Mus musculus]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase-associated	

1378	87595071 (2755, 2756)	Novel Protein sim. GBank gi4107015[dbj BAA36293] - (AB001772) PEM-5 (Ciona savignyi)			22278994, 22278998, 264093, 264094, 264259, 29331824, 29331827, 265009, 265016, 265019, 18108351, 264369, 264288, 29148627, 55811957, 264691, 18108366, 33657109, 18108368, 264635, 263981, 18108385
1379	85679344 (2757, 2758)	Novel Protein sim. GBank gi3252872 (AF035620) - BRCA1-associated protein 2 [Homo sapiens]			265020
1380	87627862 (2759, 2760)	Novel Protein sim. GBank gi4837737[gb AAD30662.1] - (AF098834) germ cell specific Y-box binding protein [Homo sapiens]	nud_rept		264510, 264512, 265009, 264288, 264564
1381	88179656 (2761, 2762)	Novel Protein sim. GBank gi4731580[gb AAD28508.1]AF12538 - (AF125384) L82A [Drosophila melanogaster]	UNCLASSIFIED		87168559, 265017, 264628, 22279002
1382	94847576 (2763, 2764)		Contains protein domain (PF00583) - Acetyltransferase (GNAT) family	UNCLASSIFIED	22278994, 22278997, 22278999, 29331822, 29331824, 29331826, 265007, 60432229, 60433356, 85658542, 265017, 265018, 264685, 284768, 21906766, 35695917, 33657023, 27486261, 27486262, 35695763, 35695855, 87168518, 22279002
1383	87860598 (2765, 2766)			UNCLASSIFIED	18108396, 264692
1384	86915895 (2767, 2768)			UNCLASSIFIED	264488, 264508, 264509, 264905, 264906, 264908, 264909, 284511, 264512, 264910, 264760, 18108351, 284768, 264769, 35695855, 264630, 264636, 264555, 264638, 264483, 264564, 264486
1385	86376788 (2769, 2770)	Novel Protein sim. GBank gi2384732 (AF015911) - NAC-1 protein [Rattus norvegicus]	UNCLASSIFIED		35696052, 55811386, 264688, 21906765, 265020, 33657023, 18108385
1386	91013049 (2771, 2772)	Novel Protein sim. GBank gi2384910 (AF022982) - contains similarity to the A-type potassium current class of channel proteins [Caenorhabditis elegans]	Inf		60432289, 29331828, 264906, 264907, 55182435, 265011, 264681, 60170615, 33657023, 83373044, 264568
1387	87797958 (2773, 2774)	Novel Protein sim. GBank gi4160304[emb CAA10600] - (AJ132192) HS1 binding protein 3 [Mus musculus]	UNCLASSIFIED		264591
1388	95101652 (2775, 2776)	Novel Protein sim. GBank gi4895184[gb AAD32753.1]AC00723 - (AC007231) putative disease resistance protein [Arabidopsis thaliana]	glycoprotein		65274572, 22278999, 264259, 29331826, 29331827, 35696052, 264509, 264907, 264908, 264909, 265006, 265008, 60170831, 33657402, 60433438, 264596, 21906754, 87168559, 284600, 265017, 264683, 18108354, 52844229, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 264692, 33657023, 33657109, 35695855, 264558, 60170394, 83373044, 22279000

1388	81256016 (2777, 2778)	Novel Protein sim. GBank gij5689387[idb]BAA82977.1] - (AB028948) KIAA1025 protein [Homo sapiens]	Contains protein domain (PF00641) - Zn-finger in Ran binding protein and others.	UNCLASSIFIED	65274572, 22278999, 264259, 28331822, 29331824, 29331825, 60432289, 29331828, 29146499, 264908, 68712502, 55812038, 265017, 265018, 265019, 18108351, 264369, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 264692, 33657023, 33657349, 18108370, 18108374, 55811576, 264555, 264556, 264557, 60170394, 83373044, 22278000, 264563, 264564, 52645156, 52646365, 264259, 52645080, 29331825, 29331826, 264906, 52644045, 285009, 33657084, 52644296, 87168474, 87168559, 265017, 265018, 264760, 264682, 264288, 264686, 264687, 56181562, 52644229, 21906765, 21906769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486261, 27486264, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1389	94111916 (2779, 2780)	Novel Protein sim. GBank gij3702295 (AC005783) - R33083_1 [Homo sapiens]	peptidase		29331826, 29331828, 29331830, 264448, 264288, 33657023, 18108365, 264555, 264556, 83373044
1391	91227345 (2781, 2782)	Novel Protein sim. GBank gij1346910[sp]P28650[PUA1_MOUSE - ADENYLOSUCINATE SYNTHETASE, MUSCLE ISOZYME (IMP--ASPARTATE LIGASE)]	Contains protein domain (PF00709) - Adenylosuccinate synthetase		52646842, 65274572, 22278994, 22278995, 35696286, 56994075, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 60432289, 29331826, 29331827, 35696052, 29331828, 33656970, 264907, 52644045, 265006, 265007, 265008, 60431735, 60433356, 52646317, 55811386, 52644296, 265010, 87168559, 265017, 264604, 265018, 265019, 264448, 264288, 264369, 264766, 21906764, 21906767, 35695917, 265020, 265021, 33657109, 52645129, 27486261, 27486262, 27486265, 33657349, 35695763, 18108370, 18108374, 18108376, 55811576, 35696423, 35695955, 264636, 52644332, 18108382, 18108385, 87168518, 60432113, 22279000, 264484, 264566, 18108391
1392	94311097 (2783, 2784)	Novel Protein sim. GBank gij726286 (U22394) - mSin3A [Mus musculus]	Contains protein domain (PF00560) - Leucine Rich Repeat	UNCLASSIFIED	264763, 264631
1393	80409472 (2785, 2786)			UNCLASSIFIED	264629
1394	15028819 (2787, 2788)			UNCLASSIFIED	265009, 18108381
1395	95361471 (2789, 2790)	Novel Protein sim. GBank gij2274845[idb]BAA21534] - (D88461) N-WASP [Rattus rattus]		UNCLASSIFIED	



1396	95363253 (2791, 2782)	Novel Protein sim. GBank gij2135904 pir  54810 - pHLEIF1 - human			22278997, 22278999, 264259, 29331825, 60432289, 29331828, 29146498, 29146499, 264907, 264908, 29331830, 264909, 265006, 265007, 265008, 265009, 60433356, 265010, 264802, 265017, 265018, 265019, 18108354, 52644228, 18108358, 21906767, 29148827, 21906768, 21906769, 29148629, 29148784, 265021, 265022, 18108368, 18108374, 56182323, 18108385, 264563, 264567, 35696286, 264907, 66712502, 264510, 35695917, 264692, 264693, 35698423, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 33656970, 87168474, 265018, 265019, 264682, 264768, 21906767, 265020, 33657023, 27486281, 55811576, 264632, 264639, 83373044, 87168518, 22278002
1397	87631317 (2783, 2794)		UNCLASSIFIED		
1398	91233667 (2785, 2786)	Novel Protein sim. GBank gij5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]			
1399	87631076 (2787, 2788)	Novel Protein sim. GBank gij2496887 sp Q09232 YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C09F5.2 IN CHROMOSOME III	UNCLASSIFIED		
1400	95418064 (2799, 2800)	Novel Protein sim. GBank gij283920 pir  S27939 - tensin - chicken	Contains protein domain (PF00017) - Src homology domain 2		56182575, 22278894, 22278897, 264259, 29331822, 29331825, 29331826, 29331827, 29331828, 264908, 56182435, 264112, 265009, 265011, 265017, 265018, 265019, 264760, 264762, 264765, 264288, 264885, 264687, 56181562, 264769, 21906768, 21906767, 55811957, 264691, 264692, 264628, 264629, 55811576, 264634, 264555, 264637, 264557, 264638, 18108381, 264558, 18108384, 60432113, 22279000
1401	91226379 (2801, 2802)	Novel Protein sim. GBank gij3256185 emb CAA15485  - (AL008635) dJ510H16.1 [Homo sapiens]	Contains protein domain (PF00790) - VHS domain	UNCLASSIFIED	55274572, 60432289, 264909, 264758, 264768, 21906769, 22278002
1402	95361475 (2803, 2804)	Novel Protein sim. GBank gij1515427 (U57523) - nel homolog [Homo sapiens]	Contains protein domain (PF00008) - EGF-like domain	Igf	264905, 264907, 264908, 264909, 264112, 264693, 33657109, 264634
1403	94147933 (2805, 2806)	Novel Protein sim. GBank gij5262615 emb CAB45747.1  - (AL080156) hypothetical protein [Homo sapiens]			65274572, 66712502, 265017, 264448, 264288, 21906765, 21906769, 264693, 55811576, 65274791, 60432113
1404	90935393 (2807, 2808)		UNCLASSIFIED		65274572, 22278998, 29331822, 29331828, 66712502, 265008, 60433438, 265017, 264693, 18108385

1405	95095068 (2809, 2810)	Novel Protein sim. GBank gij854065[embjCAA58337] - (X83413) U88 [Human herpesvirus 6]			264488, 56994075, 35696286, 29331822, 29331824, 29331826, 29331828, 35696052, 264508, 264906, 264907, 264908, 264510, 264511, 264910, 33657402, 264594, 264758, 264600, 264604, 264762, 18108351, 264764, 33657023, 33657109, 264628, 264634, 83373044, 22279002, 264563, 264482, 264486, 264567
1406	87612369 (2811, 2812)	Novel Protein sim. GBank gij624076[gjAAC96425.1] - (U42580) contains Pro-rich Px motifs: SPKPP (20X), PEPPA (9X); similar to soybean pro-rich cell wall protein, corresponds to Swiss-Prot Accession Number P13993 [Parametrium bursaria Chlorella virus 1]	collagen		264907, 264605
1407	94129872 (2813, 2814)	Novel Protein sim. GBank gij2827886 (AF015037) - endooligopeptidase A related protein; EOPA related protein [Oryctolagus cuniculus]	UNCLASSIFIED		35696286, 22278999, 264094, 264259, 66714117, 29331826, 29331827, 29331828, 29146498, 264107, 264908, 265006, 265008, 264810, 60433438, 265011, 265017, 18108351, 264448, 264288, 264686, 21906765, 21906769, 264692, 33657109, 18108370, 264628, 263972, 18108374, 35696423, 55811576, 264631, 264557, 264558, 83373044, 18108385, 87168518, 60432113, 22279002
1408	95361477 (2815, 2816)	Novel Protein sim. GBank gij2564953 (AF030001) - unknown [Mus musculus]	oncogene	Contains protein domain (PF00008) - EGF-like domain	264488, 264489, 35696286, 264109, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 265008, 265009, 264910, 33657402, 264757, 264758, 265011, 264801, 265017, 264760, 264762, 264683, 264685, 264766, 264687, 264689, 21906767, 265021, 264690, 264691, 33657023, 264692, 264693, 33657109, 264628, 264629, 35696423, 35695855, 264631, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 56182323, 264639, 264563, 264564, 264565, 264566, 264567
1409	56644385 (2817, 2818)	Novel Protein sim. GBank gij2662165[dbjBAA23714] - (AB007902) HH0712 cDNA clone for KIAA0442 has a 574-bp insertion at position 1474 of the sequence of KIAA0442. [Homo sapiens]			264693
1410	86612567 (2819, 2820)	Novel Protein sim. GBank gij2493780[spjQ6094]ACR3 MOUSE - 30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN PRECURSOR (ACRP30) (ADIPOCYTE SPECIFIC PROTEIN ADIPOQ)	complement	Contains protein domain (PF00386) - C1q domain	29331826, 264112, 264512, 265009, 265010, 264601, 264686, 264769, 21906767, 263974, 264631, 264566

1411	87818641 (2821, 2822)	Novel Protein sim. GBank gij3123155ipf91343jYM3M_CAEEL - HYPOTHETICAL 49.0 KD TRP-ASP REPEATS CONTAINING PROTEIN F55F8.5 IN CHROMOSOME 1	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinasereceptor	22278995, 22278997, 22278999, 29331822, 60432289, 29331828, 264907, 265017, 265018, 264682, 21906767, 21906768, 21906769, 265020, 264690, 264691, 33657023, 33657109, 27486264, 264628, 263972, 264634, 264558, 18108385
1412	84390919 (2823, 2824)			UNCLASSIFIED	264737
1413	95416559 (2825, 2826)	Novel Protein sim. GBank gij3879121emb CAA94370  - (Z70310) predicted using GeneFinder; Similarity to Mouse ankyrin (PIR Acc. No. S3771); cDNA EST EMBL:U01923 comes from this gene; cDNA EST EMBL:D32335 comes from this gene; cDNA EST EMBL:D32723 comes from this gene; cDNA EST E...	Contains protein domain (PF00023) - Ank repeat	homeobox	59994075, 29331822, 35696052, 29331828, 29331830, 264909, 52644045, 264510, 52644296, 85695942, 87168474, 265017, 265018, 264681, 264687, 21906768, 35695917, 265020, 52644150, 264692, 263967, 27486264, 35695763, 264639, 18108387, 264566
1414	94675860 (2827, 2828)	Novel Protein sim. GBank gij3252981 (AF068892) - Ras-binding protein SUR-8 [Mus musculus]	Contains protein domain (PF00560) - Leucine Rich Repeat	UNCLASSIFIED	264682, 264683, 265022, 264636
1415	94326948 (2829, 2830)	Novel Protein sim. GBank gij1871187 (U90439) - unknown protein [Arabidopsis thaliana]			52646365, 56182575, 22278994, 22278995, 56994075, 22278998, 22278997, 22278998, 22278998, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 29146498, 66712502, 29331830, 52644045, 264113, 264511, 33657402, 264757, 21908754, 55811386, 265017, 265018, 265019, 264761, 264683, 264369, 264288, 264686, 264689, 21906766, 21906767, 29148627, 21906769, 55811957, 265020, 265021, 264690, 33657023, 65274620, 52645129, 27486262, 27486264, 60431528, 264629, 35695855, 56182323, 264559, 60432113, 264404, 22279002, 264482

1416	84325977 (2831, 2832)	Novel Protein sim. GBank gi 510657 gb AAD39749.1 AF12305 - (AF123052) MLL sepin-like fusion protein [Homo sapiens]	Contains protein domain (PF00735) - Cell division protein	- struct	18108392, 18108394, 18108397, 18108398, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 29331822, 35696052, 29331828, 29146498, 264905, 264907, 264908, 264828, 264909, 264113, 265006, 265007, 265008, 265009, 60170831, 264595, 18108348, 21906754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264762, 18108351, 264681, 264763, 264682, 264683, 264766, 52644229, 264688, 264689, 21906765, 21906766, 21906767, 29148627, 21906788, 55811957, 29148629, 265020, 52644150, 18108361, 33657023, 18108362, 18108368, 264628, 18108370, 264629, 18108374, 18108378, 55811576, 65274791, 264634, 264636, 56182323, 18108381, 60170394, 18108385, 56526486, 87168518, 22279000
1417	87826663 (2833, 2834)	Novel Protein sim. GBank gi 4958935 dbj BAA78095.1  - (AB027570) suppressor of potassium transport defect 3 (Rattus norvegicus)		ATPase_associated	264107, 264448
1418	87594276 (2835, 2836)			UNCLASSIFIED	264259, 264908, 265010, 52644229, 21906764, 21906768, 264690, 264639, 18108388
1419	87757168 (2837, 2838)	Novel Protein sim. GBank gi 2072294 (U95097) - mitotic phosphoprotein 43 [Xenopus laevis]		struct	264259, 60432289, 265006, 87168474, 264288
1420	87298628 (2839, 2840)	Novel Protein sim. GBank		ATPase_associated	29331824, 265007, 264563
1421	94746986 (2841, 2842)	gi 5174421 ref NP_006023.1 pCPNE - copine VI (neuronal) Novel Protein sim. GBank gi 3876090 emb CAA83459.1  - (Z69635) Similarity to Yeast uridine kinase (SW:URK1_YEAST); cDNA EST EMBL:Z14695 comes from this gene; cDNA EST CEMSE17F comes from this gene; cDNA EST EMBL:D67355 comes from this gene; cDNA EST yk209h1.5 comes from this gene...		kinase	18108358, 18108396, 18108397, 21906766, 18108398, 21906767, 56182575, 21906768, 21906769, 56181686, 55811957, 35695917, 35696286, 22278996, 22278997, 22278998, 22278999, 265021, 265022, 60170615, 264259, 33657023, 29331822, 56182181, 29331824, 66714117, 29331825, 33657109, 29331826, 27486281, 29331828, 35696052, 33657349, 264905, 264509, 20281149, 18108370, 264907, 60431528, 66712502, 263972, 55811576, 35696423, 35695855, 264512, 265007, 60431850, 60432229, 60431735, 56182323, 264558, 60170394, 83373044, 55812038, 264758, 18108385, 21906754, 55811386, 87168518, 87168559, 60432113, 265017, 265018, 265019, 22279002, 55811150, 264563, 264682, 264763, 264448, 264566, 264486, 18108391

1422	88178777 (2843, 2844)	Novel Protein sim. GBank gi 4505939 ref NP_000928.1 pPOLR - polymerase (RNA) II (DNA directed) polypeptide A (220kD)		mapolymerase	56994075, 35696286, 87168559, 55811957, 55811576, 264555, 264557, 87168518
1423	86997762 (2845, 2846)			UNCLASSIFIED	264686, 264489, 264692, 264594, 264603, 265018, 264908
1424	95201610 (2847, 2848)	Novel Protein sim. GBank gi 437181 (U02289) - GTPase- activating protein [Caenorhabditis elegans]	Contains protein domain (PF00620) - struct RhoGAP domain		29331822, 29331825, 29331827, 29146498, 264905, 264906, 264908, 264909, 265007, 264910, 265009, 33109954, 265010, 87168559, 265019, 264766, 264687, 21906765, 21906766, 21906767, 21906768, 29148627, 55811957, 29148629, 265021, 264691, 264692, 56526486, 22279002, 264583
1425	21662314 (2849, 2850)	Novel Protein sim. GBank gi 100798 pir S14959 - proline- rich protein - wheat		UNCLASSIFIED	265007, 264558
1426	94322115 (2851, 2852)	Novel Protein sim. GBank gi 2078441 (U56964) - weak similarity to S. cerevisiae intracellular protein transport protein US11 (SP-P25386) [Caenorhabditis elegans]		UNCLASSIFIED	264488, 60424179, 35696286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331827, 56182435, 264910, 60433356, 60433438, 21906754, 265018, 264288, 21906765, 21906766, 21906767, 21906769, 265020, 265022, 33657109, 18108370, 18108376, 264558, 83373044, 18108385, 56526486, 22279002, 264482
1427	91227510 (2853, 2854)	Novel Protein sim. GBank gi 5916074 gb AAD4-5616.1 AF06194 - (AF061943) protate- derived STE20-like kinase PSK [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	22278994, 56994075, 22278997, 29331828, 29331830, 264828, 265006, 265007, 265008, 265009, 264683, 264288, 18108354, 21906765, 21906768, 29148629, 33657023, 18108374, 35695855, 83373044, 22278002, 264564
1428	94323008 (2855, 2856)	Novel Protein sim. GBank gi 138350 sp P28968 VGLX_HSVB - GLYCOPROTEIN X PRECURSOR		glycoprotein	56181686, 264259, 264907, 265007, 265009, 264595, 265010, 264688, 65274620, 264629, 65274791, 22279002, 264566
1429	87886689 (2857, 2858)			UNCLASSIFIED	264112, 264595, 265017, 265019, 21906765, 263977, 264555
1430	94735021 (2859, 2860)	Novel Protein sim. GBank gi 1181619 dbj BAA11565  - (D82384) a variant of TSC-22 [Gallus gallus]			264094, 29331824, 264591, 264593, 265018, 264681, 21906765, 21906767, 65274620, 55811576, 264639, 87168518, 22279002
1431	80429081 (2861, 2862)	Novel Protein sim. GBank gi 5420389 emb CAB46860.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264768, 265020, 264691, 264693, 264638
1432	87463004 (2863, 2864)	Novel Protein sim. GBank gi 414797 (L18966) - pyruvate dehydrogenase phosphatase [Bos taurus]		phosphatase	18108394, 29146498, 265007, 60433438, 264763, 29148629, 263969
1433	87605403 (2865, 2866)	Novel Protein sim. GBank gi 2460316 (AF022147) - uterus- ovary specific putative transmembrane protein [Rattus norvegicus]	Contains protein domain (PF00100) - Zona pellucida-like domain	UNCLASSIFIED	264259, 264510, 264591, 264603, 264565
1434	85713730 (2867, 2868)			UNCLASSIFIED	264682, 264691

1435	84708213 (2869, 2870)	Novel Protein sim. GBank gi 3970850 dbj BAA34789.1  - (AB015330) HRIHFB2007 [Homo sapiens]		transcripti factor	22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 66714117, 29331826, 33656970, 264508, 264905, 66712502, 29331830, 264909, 265007, 265008, 264910, 265009, 60433356, 60433438, 264596, 21906754, 265010, 265017, 265018, 265019, 18108351, 264762, 264448, 264288, 264769, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264690, 264691, 33657109, 264628, 18108374, 18108376, 55811576, 264636, 60170394, 56182323, 264559, 83373044, 87168518, 60432113, 22279000, 22279002, 264563, 264482, 264565
1436	86635024 (2871, 2872)	Novel Protein sim. GBank gi 318397 emb CAA39515  - (X56044) protein HIFC [Mus musculus]		UNCLASSIFIED	263978, 264557, 264559
1437	87631082 (2873, 2874)	Novel Protein sim. GBank gi 2496887 sp Q09232 YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C09F5.2 IN CHROMOSOME III		UNCLASSIFIED	22278997, 66714117, 29331826, 264907, 56182435, 265009, 18108351, 264692, 264693
1438	85544280 (2875, 2876)	Novel Protein sim. GBank gi 1805808 (AD000092) - hypothetical human serine-threonine protein kinase R31240_1 [Homo sapiens]	Contains protein domain (PF005595) - PDZ domain (Also known as DHR or GLGF).	kinase	264488, 264508, 264906, 264909, 264757, 264600, 264601, 264605, 264768, 264769, 264890, 35698423, 264558, 264563, 264566
1439	91231894 (2877, 2878)	Novel Protein sim. GBank gi 3876299 emb CAA94892  - (Z71180) similar to BPTKUNITZ inhibitor domain; cDNA EST EMBL:D68293 comes from this gene; cDNA EST yk448h4.5 comes from this gene; cDNA EST yk249e6.5 comes from this gene; cDNA EST yk448h4.3 comes from this gene [Caenorhabdi...	Contains protein domain (PF00450) - Serine carboxypeptidase	cathepsin	22278994, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 264107, 264508, 264509, 264907, 66712502, 29331830, 56182435, 264511, 265006, 265007, 265009, 60432229, 60433438, 264595, 55812038, 55811386, 265011, 265017, 265018, 265019, 18108351, 264448, 18108354, 264288, 18108355, 264767, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 33657109, 18108370, 18108374, 55810764, 35695855, 264634, 264638, 56182323, 83373044, 18108387, 87168518, 60432113, 22279000, 264486
1440	87423643 (2879, 2880)	Novel Protein sim. GBank gi 2662165 dbj BAA23714  - (AB007902) HH0712 cDNA clone for KIAA0442 has a 574-bp insertion at position 1474 of the sequence of KIAA0442. [Homo sapiens]		UNCLASSIFIED	264887, 264259, 264906, 264907, 264908, 264909, 265008, 265010, 265017, 265018, 265019, 18108351, 264369, 265020, 33657023, 33657109, 60431528, 55811576, 264635

1441	95317662 (2881, 2882)	Novel Protein sim. GBank glij4493959[embjCAB11123.2] - (Z98551) predicted using hexExon; MAL3P6.28 (PFC0845c). Hypothetical protein, len: 167 aa; Similarity to model organism hypothetical proteins (C.elegans, D.melanogaster, S.cerevisiae & S.pombe). C.elegans protein ZK287.5 (TR:...	Contains protein domain (PF00646) - helicase F-box domain.	18108392, 264488, 263994, 264489, 56182575, 22278994, 22278995, 56994075, 35696286, 22278997, 22278999, 60432049, 264259, 29331822, 29331824, 56182181, 66714117, 29331826, 29331827, 29331828, 35696052, 29146498, 264508, 264905, 264509, 264908, 264907, 264908, 264909, 56182435, 264510, 264511, 264512, 264910, 264592, 264593, 33657402, 60433438, 264595, 264758, 21906754, 85658542, 87168474, 265010, 87168559, 264600, 264602, 265017, 264604, 265018, 264605, 265019, 264760, 264761, 264762, 264681, 264448, 264764, 264683, 264288, 264766, 264768, 264769, 52644229, 264689, 21906765, 21805766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 52644150, 264691, 264692, 33657023, 65274620, 33637109, 35695763, 264628, 18108370, 264629, 18108374, 55811576, 35696423, 65274791, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 56182323, 264558, 60170394, 264639, 264559, 83373044, 18108385, 18108388, 56526486, 22279000, 22279002, 264563, 264483, 264564, 264566, 264567, 264488
1442	83367491 (2883, 2884)	Novel Protein sim. GBank glij5103027[dbjBAA78765.1] - (AB023419) mSox7 [Mus musculus]	transcript factor	264906, 265007, 264693, 264558
1443	87109935 (2885, 2886)	Novel Protein sim. GBank glij4887229[gbjAAD32244.1]AF15075 - (AF150755) microtubule-actin crosslinking factor [Mus musculus]	Contains protein domain (PF00435) - spectrin repeat	52645080, 264691, 264628, 264555
1444	87620478 (2887, 2888)	Novel Protein sim. GBank glij3874447[embjCAB02772] - (Z81039) predicted using Genetinder; cDNA EST EMBL: T01209 comes from this gene; cDNA EST yk278a11.3 comes from this gene; cDNA EST yk308a9.3 comes from this gene; cDNA EST yk308a9.5 com... (AJ222968) L-periaxin [Mus musculus]	UNCLASSIFIED	264259, 29331822, 29331824, 66714117, 29331828, 264288, 35695917, 33657023, 264635, 60170394
1445	84990470 (2889, 2890)	Novel Protein sim. GBank glij2959886[embjCAA11022] - (AJ222968) L-periaxin [Mus musculus]	UNCLASSIFIED	264369
1446	85079268 (2891, 2892)	Novel Protein sim. GBank glij5081610[gbjAAD39464.1]AF13544 - (AF135440) huntington yeast partner C [Mus musculus]	Contains protein domain (PF01846) - FF domain	264369
1447	86945392 (2893, 2894)	Novel Protein sim. GBank glij3980411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]	Contains protein domain (PF00439) - Bromodomain	18108336, 35696286, 22278997, 66714117, 29331828, 265009, 264758, 265018, 264288, 21906766, 21906767, 264692, 264634, 264566
1448	94990477 (2895, 2896)	Novel Protein sim. GBank glij3980411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]	Contains protein domain (PF00439) - Bromodomain	29331827, 264509, 264909, 265008, 264595, 18108357, 18108385, 264566, 264486

1449	87860859 (2897, 2898)				UNCLASSIFIED	66714117, 264908, 264908, 264591, 264601, 264764, 264632
1450	87458696 (2899, 2900)	Novel Protein sim. GBank glj1707074 (U80450) - M01E11.2 [Caenorhabditis elegans]			UNCLASSIFIED	35696286, 35696052, 265008, 265009, 60170831, 33109954, 264683, 264689, 35696423, 35695855, 56526486
1451	87787870 (2901, 2902)	Novel Protein sim. GBank glj4160304[embjCAA10600] - (AJ132192) HS1 binding protein 3 [Mus musculus]			UNCLASSIFIED	29331826, 264683, 264693, 263978, 264630
1452	85682899 (2903, 2904)	Novel Protein sim. GBank glj2832906[dbjBAA24608.1] - (D89340) dipeptidyl peptidase III [Rattus norvegicus]			peptidase	264681, 33657023, 264829
1453	86130434 (2905, 2906)	Novel Protein sim. GBank glj728831[spP39189]ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII			kinase	264510, 264768
1454	11204698 (2907, 2908)					264556
1455	87787896 (2909, 2910)				UNCLASSIFIED	29331822, 66714117, 29331825, 264905, 29331830, 265006, 265008, 265009, 265011, 265019, 18108351, 21906768, 33657109, 18108376, 264632, 56182323, 87168518
1456	86320218 (2911, 2912)	Novel Protein sim. GBank glj729230[spP41004]CUT3_SCHPO - CHROMOSOME SEGREGATION PROTEIN CUT3			transport	22278995, 22278996, 22278997, 22278998, 22278999, 29331827, 264107, 265017, 21906765, 21906766, 21906767, 21906769, 29148629, 18108370, 22279000
1457	80076900 (2913, 2914)				UNCLASSIFIED	264107, 264566
1458	87800460 (2915, 2916)	Novel Protein sim. GBank glj2246532 (U93872) - ORF 73. contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]			transport	56182575, 22278999, 60432049, 264259, 29331826, 29331827, 29331828, 264102, 264107, 264110, 265009, 60432229, 265019, 265020, 263972, 263976, 264635, 22279002, 264566
1459	95360920 (2917, 2918)	Novel Protein sim. GBank glj5524667[gbjAAD44333.1]AF15935 - (AF159356) Munc13-4 protein [Rattus norvegicus]	Contains protein domain (PF00168) - C2 domain		kinase	22278997, 264259, 29331824, 29331826, 29331827, 29331828, 265017, 265018, 264760, 264682, 264448, 264288, 264766, 265021, 264692, 33657023, 33657109, 35695855, 264566
1460	95354602 (2918, 2920)				UNCLASSIFIED	29331822, 264591, 55811957, 264691, 264693, 65274620
1461	94741513 (2921, 2922)	Novel Protein sim. GBank glj1707274 (U80931) - strong similarity to class-III of pyridoxal-phosphate-dependent aminotransferases [Caenorhabditis elegans]	Contains protein domain (PF00202) - gaba Aminotransferases class-III pyridoxal phosphate		gaba	22278997, 29331822, 35696052, 265009, 264758, 265017, 265018, 265019, 264760, 264369, 264687, 21906765, 21906768, 265022, 33657109, 27486261, 284555, 83373044
1462	87732018 (2923, 2924)				UNCLASSIFIED	264555, 264556
1463	86090605 (2925, 2926)	Novel Protein sim. GBank glj1770466[embjCAA66912] - (X98259) M-phase phosphoprotein 8 [Homo sapiens]	Contains protein domain (PF00385) - 'chromo' (CHR)romatin Organization MOffier) domain		struct	60432049, 264259, 29146499, 264906, 264907, 264512, 265017, 264763, 264766, 18108370, 18108374, 264636, 18108385, 18108388



1464	87620482 (2927, 2928)	Novel Protein sim. GBank gi 3874447 emb CAB02772  - (Z81039) predicted using GeneFinder; cDNA EST EMBL: T01209 comes from this gene; cDNA EST yk278a11.3 comes from this gene; cDNA EST yk278a11.5 comes from this gene; cDNA EST yk308a9.3 comes from this gene; cDNA EST yk308a9.5 com...		UNCLASSIFIED	264569, 22278995, 22278996, 22278997, 22278998, 29331822, 29331824, 29331825, 35696052, 20281100, 264905, 29331830, 264909, 265007, 33657402, 21906754, 265017, 265018, 264682, 264684, 264369, 264288, 264766, 21906765, 21906766, 21906767, 21906769, 35695917, 264691, 33657023, 264692, 35696423, 35695855, 264630, 264631, 264639, 264665
1465	87425192 (2929, 2930)	Novel Protein sim. GBank gi 4589598 dbj BAA76821.1  - (AB023194) KIAA0977 protein [Homo sapiens]		glucoamylase	264468, 22278994, 56994075, 60432049, 264259, 56182181, 60432289, 29331827, 52644045, 264511, 265007, 265008, 264596, 55812038, 55811386, 264600, 264602, 265017, 265018, 264604, 265019, 18108351, 18108354, 56181562, 21906769, 265021, 33657023, 33657182, 55811576, 264557, 18108382, 60432113
1466	87606227 (2931, 2932)			UNCLASSIFIED	264512, 265017, 264689, 264558
1467	87614328 (2933, 2934)	Novel Protein sim. GBank gi 2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		struct	264683, 264636
1468	95342882 (2935, 2936)	Novel Protein sim. GBank gi 4507241 ref NP_003137.1 pSSRP - structure specific recognition protein 1			22278998, 264756, 265018, 265019, 21906769, 265020, 33657109, 22279002
1469	78236174 (2937, 2938)	Novel Protein sim. GBank gi 1906586 (U81788) - kinesin-73 [Drosophila melanogaster]		struct	18108394, 18108397, 18108398, 35696052, 29146499, 265007, 265008, 285009, 265010, 265011, 18108354, 18108365, 18108368, 18108374, 18108381, 18108382, 18108384, 18108388
1470	84980482 (2939, 2940)	Novel Protein sim. GBank gi 5649170 gb AAD43131.2 AF159092 (sylv709613 protein [Homo sapiens])		UNCLASSIFIED	18108394, 18108398, 56182575, 264259, 29331822, 29331824, 29331825, 60432289, 264907, 264909, 265007, 264910, 285008, 264591, 60432229, 60433356, 264595, 60433438, 264758, 33109954, 265010, 265011, 265018, 264760, 264448, 264764, 264288, 264369, 18108357, 264769, 18108358, 21906767, 21906769, 55811957, 265021, 18108361, 264691, 18108362, 18108385, 18108368, 264628, 18108379, 264637, 264557, 18108381, 56182323, 18108382, 83373044, 18108384, 18108388, 87168518, 60432113, 264404, 22279002, 264482, 264587, 264487
1471	87826842 (2941, 2942)	Novel Protein sim. GBank gi 3876146 emb CAB01750  - (Z78542) similar to Mitochondrial carrier proteins; cDNA EST EMBL: T01651 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	- transport	29331822, 29331824, 29331825, 264828, 264603, 264689, 264693, 18108374, 55811576

1472	87756616 (2943, 2944)	Novel Protein sim. GBank gi 4680707 gb AAD27743.1 AF13296 - (AF132968) CGI-34 protein [Homo sapiens]		UNCLASSIFIED	264905, 18108351, 21906765, 264486
1473	87791609 (2845, 2946)	Novel Protein sim. GBank gi 3688780 (AF042180) - testis- specific Y-encoded-like protein [Mus musculus]	Contains protein domain (PF00956) - Nucleosome assembly protein (NAP)	MHC	18108394, 22278995, 56994075, 22278999, 29331822, 29331824, 68714117, 29331825, 29331826, 35696052, 264906, 264907, 56182435, 265007, 264758, 265018, 265019, 264760, 264764, 264288, 264685, 264686, 264768, 21906769, 55811957, 265021, 264691, 264693, 264629, 55811576, 264634, 264638, 56182323, 22279002, 264566, 264486
1474	85800989 (2947, 2948)	Novel Protein sim. GBank gi 2494890 sp Q92176 CORO_BOVIN - CORONIN-LIKE PROTEIN P57	Contains protein domain (PF00400) - WD domain, G-beta repeat	struct	264488, 35695917, 35696286, 264692, 33657023, 264693, 33657109, 35696052, 264508, 264905, 264906, 264907, 264629, 264908, 264909, 35696423, 35695855, 264511, 264910, 264632, 264634, 264635, 264636, 264637, 264556, 264557, 264639, 264758, 60432113, 264804, 264605, 264565, 264568, 264764, 264488, 264685, 264766, 264681, 264682, 264288, 264566
1475	86871935 (2949, 2950)		Contains protein domain (PF00041) - Fibronectin type III domain	UNCLASSIFIED	
1476	87548855 (2951, 2952)	Novel Protein sim. GBank gi 4757752 ref NP_004664.1 pANGP - angiopoietin 3	Contains protein domain (PF00147) - Fibrinogen beta and gamma chains, C-terminal globular domain	glycoprotein	60424179, 56181686, 29331824, 60424269, 29331826, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 264512, 265007, 265008, 265009, 264910, 33657402, 264595, 264596, 55812038, 265011, 264601, 264762, 18108351, 264288, 264368, 264685, 264766, 264689, 55811957, 264691, 264692, 264693, 18108370, 60431528, 18108374, 35696423, 264634, 264635, 264636, 60431850, 264555, 264638, 264557, 264639, 18108382, 18108388, 60432113, 22279002, 264259, 264107, 264905, 265008, 265010, 265011, 264682, 264288, 265020, 265021, 263974
1477	8774279 (2953, 2954)	Novel Protein sim. GBank gi 2498308 sp Q60870 DP1_MOUSE - POLYPOSIS LOCUS PROTEIN 1 HOMOLOG (TB2 PROTEIN HOMOLOG) (GP106)		UNCLASSIFIED	
1478	11754412 (2955, 2956)				264686

1479	91640140 (2957, 2958)	Novel Protein sim. GBank gi 5499741 gb AAD43978.1 AF15296 - (AF152961) chromatin-specific transcription elongation factor FACT 140 kDa subunit [Homo sapiens]		peptidase	56182575, 22278995, 22278996, 22278998, 22278999, 29331822, 29331824, 66714117, 264906, 264907, 56182435, 265006, 60170831, 33657402, 264758, 33109954, 21906754, 265017, 265019, 264448, 264288, 264767, 264687, 52644229, 21906764, 264689, 21906765, 21906768, 21906769, 265020, 265021, 60170615, 264691, 33657023, 33657109, 33657182, 27486261, 27486282, 33657349, 18108370, 60431528, 263976, 55811576, 264556, 264557, 60170394, 87168518, 264404, 22279000, 22279002, 264563, 264482
1480	94312412 (2959, 2960)	Novel Protein sim. GBank gi 3550456 emb CAA06329.1  - (AJ005073) Alix [Mus musculus]		UNCLASSIFIED	18108394, 65274572, 56182575, 22278995, 35696286, 56994075, 22278996, 22278997, 22278998, 22278999, 264091, 264259, 35696052, 29146499, 264103, 264105, 264108, 264907, 52644045, 264112, 265007, 265008, 265009, 60433356, 60433438, 264598, 33109954, 33657084, 52644298, 87168474, 265010, 87168559, 265017, 265018, 265019, 264448, 264682, 264683, 264769, 21906765, 21906768, 21906767, 21906768, 21906769, 265020, 265021, 60170615, 52644150, 33657109, 33657182, 263972, 35695855, 264557, 263981, 83373044, 18108385, 87168518, 264566, 264089, 264486
1481	87021442 (2961, 2962)	Novel Protein sim. GBank gi 4836807 gb AAD30566.1 AF14679 - (AF146793) PFT27 [Mus musculus]		MHC	265006, 265007, 265010, 18108374
1482	85320442 (2963, 2964)	Novel Protein sim. GBank gi 4585372 gb AAD25403.1 AF12292 - (AF122923) Wnt inhibitory factor-1 [Mus musculus]	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264908, 264910, 264758
1483	94115503 (2965, 2966)	Novel Protein sim. GBank gi 535428 (U13736) - calmodulin- like protein [Pisum sativum]	Contains protein domain (PF00036) - EF hand	- struct	264259, 29331822, 52645080, 29331825, 29331826, 33656870, 29331830, 265007, 55812038, 33109954, 265017, 264288, 21906768, 21906769, 264636, 18108380, 87168518, 22279000
1484	94131544 (2967, 2968)	Novel Protein sim. GBank gi 1911774 bbs 180090 - (S83364) putative Rab5-interacting protein (clone L1-57) [human, HeLa cells, Peptide Partial, 122 aa] [Homo sapiens]		UNCLASSIFIED	264489, 35696288, 264259, 264107, 264909, 265008, 60433356, 33657402, 60433438, 264288, 21906765, 21906766, 29146627, 33657023, 27486262, 18108374, 35696423, 83373044, 60432113
1485	80194441 (2969, 2970)	Novel Protein sim. GBank gi 5360129 gb AAD42883.1 AF15511 - (AF155117) NY-REN 62 antigen [Homo sapiens]	Contains protein domain (PF00225) - Kinesin motor domain	- struct	264369, 265020, 18108374

1486	94125066 (2971, 2972)	Novel Protein sim. GBank gij4589516[dbj]BAA76780.1] - (AB023153) KIAA0936 protein [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	56182575, 22278999, 264906, 264907, 21906754, 87168474, 265017, 265019, 18108351, 264288, 265020, 264566, 21906754, 264486
1487	86452711 (2973, 2974)	Novel Protein sim. GBank gij5019275[embj]CAB44431.1] - (AJ132751) xenobiotic/medium-chain fatty acid:CoA ligase form XL-III [Bos taurus]		synthase	
1488	87732026 (2975, 2976)	Novel Protein sim. GBank gij5712131[gbj]AAD47379.1[AF12049] - (AF120499) DEM1 protein [Homo sapiens]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	tgf	264686, 264769, 264689, 264692, 264693, 264509, 264906, 264907, 18108370, 264908, 264629, 264909, 264510, 265006, 264512, 265007, 265008, 265009, 264555, 264556, 264557, 264558, 264762, 264564, 264682, 21906767, 22278999, 265022, 264259, 264693, 29331824, 29331825, 29331826, 29331827, 29331828, 264103, 263972, 66712502, 35896423, 35895855, 265007, 265008, 265009, 83373044, 21906754, 56526486, 265017, 264563, 18108351, 284564, 264588, 264369, 264288
1489	95104277 (2977, 2978)	Novel Protein sim. GBank gij2497303[sp]Q62786[FP]PRP_RAT - PROSTAGLANDIN F2-ALPHA RECEPTOR REGULATORY PROTEIN PRECURSOR (PROSTAGLANDIN F2-ALPHA RECEPTOR ASSOCIATED PROTEIN)	Contains protein domain (PF00047) - Immunoglobulin domain	prostaglandin	21906767, 22278999, 265022, 264259, 264693, 29331824, 29331825, 29331826, 29331827, 29331828, 264103, 263972, 66712502, 35896423, 35895855, 265007, 265008, 265009, 83373044, 21906754, 56526486, 265017, 264563, 18108351, 284564, 264588, 264369, 264288
1490	87390127 (2979, 2980)			UNCLASSIFIED	56182575, 264259, 29331822, 29331824, 66714117, 29331827, 29331828, 264508, 264905, 66712502, 265007, 265008, 264594, 33657402, 55812038, 87168474, 265018, 18108351, 264369, 264288, 264769, 264689, 21906767, 21906768, 55811957, 60170615, 33657109, 35895855, 264635, 60170394, 56526486, 22279002, 264563, 265007, 264448, 18108372, 264558, 56182323
1491	83594305 (2981, 2982)	Novel Protein sim. GBank gij295671 (L11275) - selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III [Saccharomyces cerevisiae]		UNCLASSIFIED	
1492	85805363 (2983, 2984)	Novel Protein sim. GBank gij1656005 (U71205) - rit [Mus musculus]	Contains protein domain (PF00071) - Ras family	oncogene	22278997, 22278998, 29331822, 264907, 66712502

1493	81677215 (2885, 2986)	Novel Protein sim. GBank gij5689515jdbj BAA83041.1  - (AB029012) KIAA1089 protein [Homo sapiens]		UNCLASSIFIED	264488, 52646365, 65274572, 56182575, 22278994, 35696286, 58994075, 22278999, 60432049, 29331824, 29331828, 35696052, 264508, 264908, 264908, 52644045, 264909, 56182435, 265006, 265008, 265009, 60170831, 33657402, 55812038, 265010, 265011, 265017, 265018, 265019, 55811150, 264448, 264682, 264685, 264686, 52644229, 21906765, 21906766, 21908768, 21908769, 265020, 265021, 60170615, 52644150, 33657023, 18108384, 18108365, 33657109, 33657182, 27486261, 27486262, 27486264, 33657349, 27486265, 35695763, 18108370, 264629, 18108374, 52644332, 56182323, 87168518, 22279002, 264564, 264566, 264567
1484	87605265 (2987, 2988)	Novel Protein sim. GBank gij728832sp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII	Contains protein domain (PF01352) - KRAB box	kinase	264907, 265009, 264768, 18108370, 55811576, 264639, 264565, 264488
1485	87605267 (2989, 2990)	Novel Protein sim. GBank gij458958jdbj BAA76816.1  - (AB023189) KIAA0972 protein [Homo sapiens]	Contains protein domain (PF01352) - KRAB box	transcriptfactor	22278997, 264259, 264806, 264907, 265009, 264594, 33657084, 265017, 264760, 264448, 33657109, 264630, 264634, 56526486, 264563, 264565, 264566, 264486, 264567
1486	87784322 (2991, 2992)	Novel Protein sim. GBank gij5420387 emb CAB46679.1  - (A243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	35696286, 264906, 265019, 264693
1497	81695428 (2993, 2994)	Novel Protein sim. GBank gij3874925 emb CAA92591  - (Z68286) Similarity to Mouse A-RAF proto-oncogene serine/threonine-protein kinase (SW:KRAA_MOUSE); cDNA EST EMBL:D27610 comes from this gene; cDNA EST EMBL:TO1018 comes from this gene; cDNA EST EMBL:D33256 comes from this gene....	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264910, 264758, 265011, 264764, 264288, 264690, 264634, 264635, 56526486
1488	90934938 (2995, 2996)	Novel Protein sim. GBank gij728836sp P39183 ALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII		oncogene	264488, 65274572, 29331822, 265017, 265018, 21906765, 29148627, 29148629, 18108374, 264637, 264638, 264567
1499	86451589 (2997, 2998)	Novel Protein sim. GBank gij2570198 (U54556) - microfilament sheath protein SHP3 [Litomosoides sigmodontis]		glucoamylase	263978, 264566
1500	80499386 (2998, 3000)	Novel Protein sim. GBank gij2078483 (U43200) - antifreeze glycoprotein precursor [Boreogadus saida]		UNCLASSIFIED	22278999, 264769, 18108379
1501	85795297 (3001, 3002)	Novel Protein sim. GBank gij2078483 (U43200) - antifreeze glycoprotein precursor [Boreogadus saida]		UNCLASSIFIED	264559
1502	80206141 (3003, 3004)				264508, 264112, 264604, 264684, 52644150, 55811576, 264632, 264556, 264638, 56182323, 264563, 264486
1503	87012701 (3005, 3006)	Novel Protein sim. GBank gij3900855 (AC004874) - similar to N-acetylgalactosaminyltransferase; similar to Q07537 (PID:g1171889) [Homo sapiens]	Contains protein domain (PF00535) - Glycosyl transferases	transferase	29331822, 265007, 264369

1504	79540051 (3007, 3008)		Contains protein domain (PF00023) - Ankr repeat	UNCLASSIFIED	264693	29331826, 35696052, 264509, 264906, 264807, 264908, 264909, 264511, 264910, 55812038, 264759, 264763, 264764, 264689, 35695917, 265022, 33657109, 18108374, 264631, 264635, 264638, 264568
1505	86102672 (3009, 3010)	Novel Protein sim. GBank gij475375jemb[CAB41970.1] - (A132545) protein kinase [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase		65274572, 56182575, 56994075, 22278997, 22278998, 22278999, 264091, 264092, 60432049, 264259, 52645080, 29331822, 29331827, 264106, 29331830, 264908, 56182435, 264110, 264511, 264512, 55812038, 21906754, 87168559, 264600, 265017, 265018, 264681, 18108354, 264369, 264687, 21906769, 29148627, 21906768, 21906769, 29148629, 52644150, 33657023, 18108376, 65274791, 56182323, 264558, 264559, 18108385, 87168518, 60432113, 22279000, 264565
1506	94143219 (3011, 3012)	Novel Protein sim. GBank gij1304201dbj[BAA06170] - (D29768) alternatively spliced product [Rattus norvegicus]	Contains protein domain (PF00018) - SH3 domain	glycoprotein		264639
1507	83738250 (3013, 3014)	Novel Protein sim. GBank gij5689513dbj[BAA83040.1] - (AB029011) KIAA1088 protein [Homo sapiens]		helicase		264593
1508	11618758 (3015, 3016)	Novel Protein sim. GBank gij5031975jelpNP_005875.1pPAK4 - protein kinase related to S. cerevisiae STE20 effector for Cdc42Hs	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase		264259, 29331822, 29331824, 29331828, 264591, 33109954, 264563
1509	87318451 (3017, 3018)	Novel Protein sim. GBank gij113161ispIP28614JACOR_ALCEU - ACETON CATABOLISM REGULATORY PROTEIN		UNCLASSIFIED		264259, 29331822, 265007, 18108374, 264556
1510	95362643 (3019, 3020)	Novel Protein sim. GBank gij128831ispIP39188JALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY IIII		UNCLASSIFIED		264259, 29331822, 29331824, 29331827, 264508, 264907, 264510, 264511, 265007, 264512, 265008, 87168559, 264288, 265022, 33657023, 35695855, 264637, 264638, 264563
1511	88318073 (3021, 3022)	Novel Protein sim. GBank gij4559353jpbJAD23014.1AC00658 - (AC006585) putative extragenic suppressor protein [Arabidopsis thaliana]	Contains protein domain (PF01163) - RIO1/ZK632.3/MJO444 family			52645156, 18108386, 56994075, 60432289, 265006, 60433358, 60433438, 21906754, 87168474, 87168559, 265018, 264762, 264763, 264687, 21906765, 21906769, 27486262, 35695763, 18108374, 35696423, 264555, 18108385, 18108387, 18108388, 87168518, 264482
1512	95345390 (3023, 3024)	Novel Protein sim. GBank gij4559353jpbJAD23014.1AC00658 - (AC006585) putative extragenic suppressor protein [Arabidopsis thaliana]				35695052, 264905, 264906, 264907, 264908, 264909, 264910, 264591, 264766, 264689, 264692, 264629, 264636
1513	87436228 (3025, 3026)	Novel Protein sim. GBank gij1330394 (U58761) - C01F1.6 gene product [Caenorhabditis elegans]				

1514	95345392 (3027, 3028)	Novel Protein sim. GBank gi 459333 gb AAD23014.1 AC006585 putative extragenic suppressor protein [Arabidopsis thaliana]	Contains protein domain (PF01163) - RIO1/ZK632.3/MJ0444 family	UNCLASSIFIED	52644507, 52645156, 52646365, 52646842, 65274572, 22278994, 35698288, 58994075, 284259, 52645080, 29331822, 29331825, 35696052, 29331830, 52644045, 58182435, 265006, 60433358, 60433438, 55812038, 21906754, 52646317, 52644296, 87168474, 87168559, 264448, 52644229, 21906785, 21906766, 21906767, 21906768, 35695917, 265020, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486264, 27486265, 35695763, 18108376, 35698423, 35695855, 52644332, 18108385, 18108387, 87168518, 60432113 265020, 264639
1515	79163538 (3029, 3030)	Novel Protein sim. GBank gi 3879501 emb CAA87795  - (Z47812) similar to ubiquitin carboxyl-terminal hydrolase; cDNA EST EMBL:D33366 comes from this gene; cDNA EST EMBL:D33965 comes from this gene; cDNA EST EMBL:D33822 comes from this gene; cDNA EST EMBL:D34547 comes from this ge...	ubiquitin		
1516	88073539 (3031, 3032)	Novel Protein sim. GBank gi 488015 (L27479) - X123 [Homo sapiens]	UNCLASSIFIED		265008, 56182323, 22279002
1517	87793325 (3033, 3034)	Novel Protein sim. GBank gi 3415134 (AF082024) - Phyb1 [Pimpinella brachycarpa]			264081, 18108370, 264404
1518	87350697 (3035, 3036)	Novel Protein sim. GBank gi 728838 sp P39195 ALU8_HUMAN - IIII ALLU SUBFAMILY SX WARNING ENTRY IIII	Im7		66714117, 264508, 264509, 264905, 264510, 284910, 264591, 264595, 264288, 264766, 264769, 18108374, 264636, 264638, 264486 264589, 264489, 60432049, 265009, 33657402, 264596, 21906754, 265019, 264369, 21906765, 21906768, 21906769, 264691, 65274620, 33657182, 27486261, 18108374, 264557, 264639, 87168518, 22279002
1519	84328689 (3037, 3038)	Novel Protein sim. GBank gi 526268 emb CAB45771.1  - (AL080198) hypothetical protein [Homo sapiens]			
1520	87592855 (3039, 3040)	Novel Protein sim. GBank gi 2662161 dbj BA23712  - (AB007900) HH0452 cDNA clone for KIAA0440 has a 438- bp insertion at position 1711 of the sequence of KIAA0440. [Homo sapiens]	UNCLASSIFIED		18108392, 60432049, 264259, 29331824, 265007, 60433356, 265010, 21906768, 264636
1521	86970698 (3041, 3042)	Novel Protein sim. GBank gi 5052351 gb AAD38516.1 AF13542 - (AF135421) GDP- mannose pyrophosphorylase B [Homo sapiens]	Contains protein domain (PF00483) - Nucleotidyl transferase	synthase	18108394, 264259, 66714117, 265011, 264603, 265019, 18108364, 35698423, 264557, 264558, 18108388
1522	78960687 (3043, 3044)	Novel Protein sim. GBank gi 3776567 (AC005388) - Strong similarity to F21B7.33 gi 2809264 from A. thaliana BAC gb AC002560. EST gb N65119 comes from this gene. [Arabidopsis thaliana]	UNCLASSIFIED		29331824, 265018, 265020, 265021 65274572, 21906768, 264693
1523	91005151 (3045, 3046)		UNCLASSIFIED		
1524	80203723 (3047, 3048)		UNCLASSIFIED		264112, 21906754, 263974
1525	87799867 (3049, 3050)	Novel Protein sim. GBank gi 4759040 ref NP_004283.1 pRIN1 - ras inhibitor	UNCLASSIFIED		264683, 264687, 284689, 264690, 264692, 264693

1526	85105344 (3051, 3052)	Novel Protein sim. GBank gi 728850 sp P08640 JAMYH_YEAST - GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)		glycoprotein	35606286, 56182181, 60431735, 264595, 55812038, 264605, 264883, 21906765, 55811957, 265020, 65274791, 264555, 264556, 264557, 264558, 264559, 83373044
1527	85262512 (3053, 3054)	Novel Protein sim. GBank gi 2792496 (AF041107) - tulip 2 [Rattus norvegicus]			56182575, 264259, 60432049, 29331822, 60432289, 264908, 66712502, 60433438, 87168559, 265017, 264288, 21906768, 21906769, 263977, 55811576, 56182323, 18108381
1528	94130918 (3055, 3056)			UNCLASSIFIED	22278995, 22278997, 264259, 66712502, 264596, 265017, 265019, 264882, 264448, 264883, 264764, 264685, 264686, 21906765, 21906766, 21906767, 21906768, 21906769, 265022, 264693, 83373044, 18108385
1529	94120793 (3057, 3058)	Novel Protein sim. GBank gi 406663 gb AAD20053] - (AF131826) Unknown [Homo sapiens]		UNCLASSIFIED	264488, 263994, 56182575, 22278995, 35696286, 22278997, 264259, 29331822, 60432289, 29331827, 35696052, 264509, 264906, 264907, 264908, 264909, 52644045, 56182435, 264511, 265009, 264910, 60433358, 60433438, 265017, 265018, 264760, 264448, 264764, 264368, 264288, 264768, 18108357, 264768, 52644229, 21906765, 21906766, 21906767, 21906768, 265021, 265022, 52644150, 33657109, 264629, 35695855, 80432113, 22279002, 264563, 264564, 264486, 264567
1530	95012765 (3059, 3060)	Novel Protein sim. GBank gi 2828710 (AF043642) - maltrin cyclophilin [Rattus norvegicus]			264488, 264489, 35696286, 29331825, 35696052, 264508, 264905, 264906, 264907, 264909, 264510, 264511, 264512, 264810, 264592, 264595, 18108351, 264764, 264683, 264684, 264768, 264769, 18108357, 264769, 35695917, 264628, 264629, 18108374, 35695855, 264630, 264631, 264634, 264555, 264636, 264637, 264404, 264563, 264566, 264486
1531	95419351 (3061, 3062)	Novel Protein sim. GBank gi 1905874 (U90878) - carboxyl terminal LIM domain protein [Homo sapiens]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).	kinase	56182575, 35696286, 264097, 264259, 29331822, 29331825, 29331826, 29331827, 35696052, 264509, 56182435, 264510, 264511, 265007, 60433358, 55811386, 264681, 264369, 264288, 264766, 264687, 55811957, 35695917, 33657023, 35695763, 55810764, 35696423, 55811576, 263981, 60170394, 56182323, 83373044, 60432113, 264566



1532	85718224 (3063, 3064)	Novel Protein sim. GBank gij3874716[emb]CAA91265] - (Z66494) cDNA EST EMBL:D65271 comes from this gene; cDNA EST EMBL:D64845 comes from this gene; cDNA EST EMBL:D64449 comes from this gene; cDNA EST EMBL:D67438 comes from this gene; cDNA EST EMBL:D68087 comes from this gene; cDN... (Z78141) unknown [Mus musculus]		UNCLASSIFIED	284689
1533	94239830 (3065, 3066)	Novel Protein sim. GBank gij1490324[emb]CAB01543] - (Z78141) unknown [Mus musculus]	struct		29331824, 29146499, 264907, 264112, 265008, 265011, 265017, 265018, 264762, 18108351, 263967, 20281149, 18108374, 263981, 264568
1534	95343941 (3067, 3068)	Novel Protein sim. GBank gij81286[pir]S22697 - extensin - Volvox carter (fragment)		UNCLASSIFIED	264905, 264907, 264766, 264637
1535	90936732 (3069, 3070)				65274572, 22278997, 264259, 60432049, 29331822, 60432289, 29331827, 29146499, 265008, 265008, 60170831, 60433438, 33109954, 87168559, 265018, 18108357, 21906768, 29148629, 265021, 265022, 18108377, 56182323, 60432113, 22279000, 22279002
1536	87602856 (3071, 3072)	Novel Protein sim. GBank gij106024[pir]B32891 - finger protein 2, placental - human	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264686, 18108357, 18108394, 21906767, 21906768, 29148629, 35696286, 265020, 265021, 52844150, 264693, 68714117, 29331825, 29331826, 264508, 264905, 20281149, 264909, 18108374, 35696423, 35695855, 265009, 264634, 264636, 264638, 18108385, 56526488, 265017, 265018, 264563, 264762, 18108351, 264448, 264369, 264766
1537	95354556 (3073, 3074)	Novel Protein sim. GBank gij3876332[emb]CAB02086] - (Z79754) cDNA EST EMBL:T01054 comes from this gene; cDNA EST EMBL:D73600 comes from this gene; cDNA EST yk426112.5 comes from this gene; cDNA EST yk342110.5 comes from this gene; cDNA EST yk475c5.5 comes from this gene; cDNA ES...			65274572, 56182575, 60432049, 264259, 29331828, 265006, 265007, 60433356, 60433438, 264601, 18108351, 264448, 264369, 264288, 33657023, 65274620, 33657109, 60432113
1538	85724628 (3075, 3076)	Novel Protein sim. GBank gij403440 (M81787) - [Gallus domesticus skeletal muscle mRNA, partial cds.], gene product [Gallus gallus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	18108394, 18108397, 264909, 265008, 265009, 265010, 18108351, 264638, 18108382, 18108385, 18108388

1539	95337628 (3077, 3078)	Novel Protein sim. GBank gij3218411[emb CAA19575.1]-(AL023859) SPBC19C7.07c, putative RNA splicing endonuclease ga mma subunit, len:284aa, similar eg. to YAR008W, YAH8_YEAST, P39707, YAR008W, RNA splicing endonuclease gamma subunit, (275aa), fasta scores, opt:269, E(16.4e-2...	nuclease	22278994, 22278996, 35696286, 56994075, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 60424269, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 56182435, 265009, 33657402, 60433358, 60433438, 55812038, 21906754, 87168559, 265017, 265018, 265019, 18108351, 264288, 52844229, 18108359, 21906764, 21906767, 21906768, 35695917, 265020, 265021, 52844150, 33657023, 33657109, 27486261, 18108370, 18108376, 35696423, 55811576, 65274791, 284558, 56182323, 60170394, 83373044, 87168518, 60432113, 22279000, 22279002, 264566
1540	95352858 (3079, 3080)	Novel Protein sim. GBank gij5052634[gb AAD38647.1 AF14567 - (AF145672) BcDNA.GH12174 [Drosophila melanogaster]	UNCLASSIFIED	264369, 264691, 263978
1541	95317848 (3081, 3082)	Novel Protein sim. GBank gij5052349[gb AAD38515.1 AF13501 - (AF135016) protein phosphatase 2A 48 kDa regulatory subunit [Homo sapiens]	phosphatase	264488, 264489, 22278999, 264259, 29331822, 35696052, 264508, 264509, 264905, 264907, 264908, 264511, 264512, 264910, 264592, 264761, 264762, 264448, 264764, 264288, 264687, 21908769, 55811957, 35695917, 265020, 264691, 33657023, 264692, 33657109, 264628, 18108374, 264632, 264634, 264635, 264639, 18108385, 264563, 264564, 264565, 264566, 264486
1542	90937549 (3083, 3084)	Novel Protein sim. GBank gij5305702[gb AAD41779.1 AF12666 - (AF126667) calpain-like protease [Mus musculus]	cathepsin	18108392, 18108394, 65274572, 29331822, 264508, 265007, 265008, 265009, 265011, 264682, 18108354, 18108355, 52644150, 18108368, 264636, 18108381, 18108382
1543	84348768 (3085, 3086)	Novel Protein sim. GBank gij728832[sp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII]	nuclease	264106, 33109954, 265019, 264683, 35695917, 264690, 264692, 33657109
1544	87757295 (3087, 3088)	Novel Protein sim. GBank gij3493162 (AF084259) - bromodomain-containing protein BP75 [Mus musculus]	nucd_rept	35696286, 56994075, 22278999, 35696052, 60433356, 60433438, 265011, 264683, 33657109, 35696423, 264631, 87168518, 22279000
1545	85757873 (3089, 3090)	Novel Protein sim. GBank gij1086591 (U41007) - similar to S. cerevisiae nuclear protein SNF2 (SP-P22082) in a region of gly-arg repeats [Caenorhabditis elegans]	UNCLASSIFIED	264112, 264692, 264693, 55811576
1546	78476589 (3091, 3092)		UNCLASSIFIED	264905, 264686
1547	86999594 (3093, 3094)	Novel Protein sim. GBank gij2661132 (AF035683) - p21 [Mus musculus]	UNCLASSIFIED	264259, 29331822, 66714117, 265007, 55811386, 265010, 264600, 265017, 265019, 264288, 264768, 265020, 265022, 55811576, 18108380, 264563

1548	94233065 (3085, 3096)	Novel Protein sim. GBank gi 3043692 dbj BAA25510  - (AB011156) KIAA0584 protein [Homo sapiens]		UNCLASSIFIED	29331824, 60431528, 264639, 56182323
1549	95330048 (3097, 3098)	Novel Protein sim. GBank gi 5689519 dbj BAA3043.1  - (AB029014) KIAA1091 protein [Homo sapiens]		eph	60424179, 22278995, 35696286, 22278998, 22278999, 264092, 264094, 29331822, 56182181, 29331824, 35696052, 264905, 264906, 264908, 264909, 265006, 264511, 265008, 60431735, 60433356, 21906754, 55811386, 87168559, 265017, 265018, 265019, 55811150, 264682, 264288, 264369, 56181562, 264769, 21906765, 21906788, 21906769, 55811957, 265020, 264691, 33857109, 60431528, 35696423, 35695855, 56526486, 60432113, 22279002, 264563, 264566
1550	95201807 (3099, 3100)	Novel Protein sim. GBank gi 544483 sp P35350 GUSB_BOVIN - POSSIBLE GUSTATORY RECEPTOR TYPE B (PPR1 PROTEIN)	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	Im7	65274572, 60432289, 265008, 264810, 265011, 265017, 265019, 264768, 56182323
1551	88077111 (3101, 3102)	Novel Protein sim. GBank gi 4758566 ref NP_004788.1 pHS6S - heparan-sulfate 6-sulfotransferase		UNCLASSIFIED	22278999, 29331822, 264508, 264509, 264908, 264907, 264908, 265007, 264512, 264910, 21906754, 265018, 265019, 264681, 264764, 264766, 264688, 264769, 21906769, 264692, 35695763, 264635, 264555, 264556, 264557, 264638, 264558, 264563
1552	87617114 (3103, 3104)			UNCLASSIFIED	264259, 29331828, 66712502, 264764, 264288, 264686, 33657109, 264558
1553	94725512 (3105, 3106)	Novel Protein sim. GBank gi 4589570 dbj BAA76807.1  - (AB023180) KIAA0963 protein [Homo sapiens]	Contains protein domain (PF00304) - Gamma-thionins family	dehydrogenase	56182575, 35696286, 29146499, 264509, 264907, 264908, 264909, 56182435, 265006, 265008, 265009, 264910, 264757, 264758, 265017, 55811150, 18108351, 264764, 56181562, 35695917, 264693, 33657109, 18108374, 35696423, 65274791, 35695855, 264635, 264555, 56182323, 18108382, 83373044, 22279000
1554	94233069 (3107, 3108)	Novel Protein sim. GBank gi 3043692 dbj BAA25510  - (AB011156) KIAA0584 protein [Homo sapiens]	Contains protein domain (PF00446) - Gonadotropin-releasing hormones	transferase	35696286, 22278997, 264259, 29331822, 29331824, 29331825, 29331828, 265007, 265009, 60432229, 33657402, 55812038, 265011, 265019, 264681, 264369, 264686, 264767, 264768, 21906765, 21906769, 35695917, 264693, 18108370, 60431528, 55811576, 264631, 60170394, 56182323, 83373044, 18108385, 22279000, 22279002

1555	87332970 (3109, 3110)	Novel Protein sim. GBank gij2237495[dijBAA21392] - (AB004534) p015 [Schizosaccharomyces pombe]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	284259, 29331826, 35696052, 264508, 264905, 284906, 264907, 264908, 5264045, 264909, 284910, 60432229, 60433356, 55812038, 264758, 264759, 33657084, 265011, 87168559, 264601, 265018, 265019, 264763, 264764, 264288, 264766, 264768, 21906765, 35695917, 265022, 264691, 33657023, 35696423, 35695855, 264635, 264555, 264636, 264638, 264639, 18108385, 56526486
1556	91228288 (3111, 3112)			UNCLASSIFIED	83373044, 264758, 265022, 264600, 35696052, 264630, 35696423, 265018, 264632, 264682, 29331822, 265020, 265011, 60432289, 264509, 264906, 264907, 264908, 264909, 264910, 264758, 55811386, 264761, 264762, 264766, 264769, 264690, 263978, 264634, 264635, 264639, 264564, 264486, 22278994, 22278996, 22278997, 22278998, 22278999, 60432048, 264259, 29331824, 29331825, 29331826, 29331827, 264908, 264909, 60433356, 21906754, 265017, 265018, 264448, 21906767, 265021, 265022, 33657023, 33657109, 18108370, 55811576, 83373044, 87168518, 22279000, 22279002, 264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1557	87640609 (3113, 3114)	Novel Protein sim. GBank gij3328611 (AF078783) - contains similarity to C3HC4-type zinc fingers (Pfam; zfc3HC4.hmm, score; 34.08); most similar to drosophila goliath protein (SW: Q06003) [Caenorhabditis elegans]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1558	84840376 (3115, 3116)	Novel Protein sim. GBank gij5360105[gjAAD42871.1]AF155105 putative zinc finger protein NY-REN-34 antigen [Homo sapiens]			264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1559	88224865 (3117, 3118)	Novel Protein sim. GBank gij112908[spjP02750]A2GL_HUMAN - LEUCINE-RICH ALPHA-2-GLYCOPROTEIN (LRG)	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1560	84580675 (3119, 3120)	Novel Protein sim. GBank gij3880146[lemb]CAA92704] - (Z88319) Similarity to Human hnRNP F protein (PIR Acc. No. S43484); cDNA EST EMBL:D34218 comes from this gene; cDNA EST EMBL:D37248 comes from this gene; cDNA EST EMBL:D71817 comes from this gene; cDNA EST EMBL:D74531 comes from...		UNCLASSIFIED	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1561	86609159 (3121, 3122)			UNCLASSIFIED	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1562	83359682 (3123, 3124)			UNCLASSIFIED	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1563	85508694 (3125, 3126)			UNCLASSIFIED	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855
1564	87766371 (3127, 3128)	Novel Protein sim. GBank gij1168287[spjP45953]ACDV_RAT - ACYL-COA DEHYDROGENASE, VERY-LONG-CHAIN SPECIFIC PRECURSOR (VLCAD)	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	UNCLASSIFIED	264259, 29331822, 60432289, 35696052, 264107, 264110, 21906754, 33109954, 87168559, 264760, 264763, 21906764, 21906765, 21906768, 265021, 264690, 35695855

1565	87783381 (3129, 3130)	Novel Protein sim. GBank gij129726[sp]P05307[PDI_BOVIN] - PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (PROLYL 4-HYDROXYLASE BETA SUBUNIT) (CELLULAR THYROID HORMONE BINDING PROTEIN) (P55)		isomerase	264488, 264689, 18108398, 55811957, 264534, 264259, 264508, 264905, 264509, 264908, 18108372, 264510, 264511, 264512, 265008, 264630, 265009, 264910, 264635, 264638, 264591, 264555, 264592, 264637, 264593, 264594, 264595, 264596, 265011, 264603, 22279002, 18108351, 264762, 264565, 264567
1566	87424749 (3131, 3132)	Novel Protein sim. GBank gij3880445[emb]CAA20329[- (AL031266) VM106R.1 [Caenorhabditis elegans]		Inf	22278996, 22278998, 264259, 29331822, 29331824, 60432289, 29331827, 66712502, 264908, 265008, 18108351, 52644228, 21908785, 21906767, 21908768, 21908769, 33657109, 264555, 264639, 264482
1567	84999008 (3133, 3134)	Novel Protein sim. GBank gij4929699[gb]AAD34110.1[AF151873] CGI-115 protein [Homo sapiens]		UNCLASSIFIED	56182575, 21906769, 264692
1568	87648761 (3135, 3136)	Novel Protein sim. GBank gij4827063[ref]NP_005072.1[pZNF1 - zinc finger protein 142 (clone pHZ-49)]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	29331827, 29331830, 264511, 265009, 264758, 21908767, 21908768, 264691, 264693, 22279000, 22279002
1569	90936668 (3137, 3138)	Novel Protein sim. GBank gij5689451[dbj]BAAB3009.1[- (AB028980) KIAA1057 protein [Homo sapiens]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	65274572, 29331822, 29331824, 29331828, 264905, 56182435, 265007, 265019, 264764, 21908765, 21906769, 55811957, 60170615, 52644150, 264692, 33657023, 33657109, 18108377, 264563, 264587
1570	86943981 (3139, 3140)	Novel Protein sim. GBank gij1255430 (U53155) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	264595, 264682, 265021
1571	91210340 (3141, 3142)	Novel Protein sim. GBank gij4507731[ref]NP_001081.1[pTUBG - tubulin, gamma polypeptide]	Contains protein domain (PF00091) - Tubulin/Fisz family	tubulin	22278996, 35696286, 22278997, 264091, 264259, 29331824, 29331825, 29331827, 35696052, 264508, 264905, 56182435, 264510, 265007, 264758, 265011, 18108351, 264448, 264288, 264369, 21906765, 21908767, 21906768, 21908769, 35695917, 265020, 265021, 33657023, 264693, 18108370, 18108377, 35696423, 35695855, 264634, 264555, 264558, 18108384
1572	81201664 (3143, 3144)			UNCLASSIFIED	35696052, 264905, 264906, 264908, 264910, 264758, 264766, 35695917, 264637
1573	80207066 (3145, 3146)			UNCLASSIFIED	283972

1574	94216142 (3147, 3148)	Novel Protein sim. GBank gi 4758334 ref NP_004256.1 pFADS - delta-6 fatty acid desaturase		Contains protein domain (PF00173) - cytochrome heme-binding domain in cytochrome b5 and oxidoreductases	18108394, 264887, 18108397, 18108398, 22278996, 22278997, 22278999, 264259, 29331825, 29331827, 29146498, 29146499, 264107, 264907, 264909, 52644045, 264511, 265008, 264910, 265009, 264591, 21906754, 265011, 265019, 18108351, 264682, 264763, 264764, 18108354, 264389, 264288, 264685, 264768, 264686, 264768, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 29148629, 264690, 264691, 264693, 20281069, 18108370, 18108374, 18108379, 35695855, 264634, 18108384, 18108385, 22279002, 264563, 264566
1575	95340019 (3149, 3150)	Novel Protein sim. GBank gi 3881810 amb CAA94856  - (Z70783) similar to EF-hand calcium binding protein; cDNA EST EMBL:CO8700 comes from this gene [Caenorhabditis elegans]		Contains protein domain (PF00036) - EF hand	56994075, 264259, 29331822, 29331824, 29331825, 60432289, 29331828, 264909, 265008, 265008, 265009, 265010, 87168559, 55811150, 264448, 18108354, 264389, 264288, 18108357, 55811957, 265020, 265021, 60170615, 264691, 33657023, 33657109, 60431528, 65274791, 35695855, 18108385, 60432113, 22279002, 264482
1576	95314019 (3151, 3152)	Novel Protein sim. GBank gi 2773195 (AF039711) - contains similarity to Physcomitrella patens glyceraldehyde 3-phosphate dehydrogenase (GB:X72381) [Caenorhabditis elegans]		UNCLASSIFIED	264568, 264092, 264094, 264095, 264259, 264508, 264905, 264508, 264907, 264909, 264510, 264511, 265006, 264910, 21906754, 265010, 265011, 87168559, 264761, 264762, 264288, 264768, 264769, 264691, 264693, 35695855, 264632, 264634, 264635, 264638, 83373044, 264488
1577	87613800 (3153, 3154)	Novel Protein sim. GBank gi 2499130 sp P70315 WASP_MOUSE - WISKOTT- ALDRICH SYNDROME PROTEIN HOMOLOG (WASP)		UNCLASSIFIED	264511, 265011, 264681, 264369, 264686, 264689, 264629, 264555, 264558, 264559
1578	87123138 (3155, 3156)				264259, 29331826, 265017, 264689, 264693, 60432113
1579	88085141 (3157, 3158)	Novel Protein sim. GBank gi 2978255 dbj BAA25190  - (AB007407) myeloid zinc finger protein-2 [Mus musculus]		Contains protein domain (PF00096) - Zinc finger, C2H2 type	35696286, 264908, 264909, 60433438, 55811386, 264369, 264685, 33657023, 264555, 264558, 264557, 87188518
1580	87255702 (3159, 3160)	Novel Protein sim. GBank gi 4324682 gb AA016986  - (AF109674) late gestation lung protein 1 [Rattus norvegicus]		Contains protein domain (PF00186) - SCP-like extracellular protein	22278999, 35696052, 29331830, 52644045, 55812038, 87168474, 265018, 264448, 265022, 264638, 56526486, 22279000
1581	95087431 (3161, 3162)	Novel Protein sim. GBank gi 2088838 (AF003386) - F59E12.4 gene product [Caenorhabditis elegans]			22278995, 29331822, 29331824, 29331826, 56182435, 264595, 55812038, 87168559, 265017, 264288, 21906764, 55811957, 35695917, 264692, 55811576, 264637, 56182323, 264559, 83373044, 60432113

1582	95358052 (3163, 3164)	Novel Protein sim. GBank gij5420387[emb CAB46679.1]- (AJ243459) proteophosphoglycan [Leishmania major]		phosphatase	264259, 604322289, 29331827, 264509, 264905, 264906, 264907, 264909, 264910, 264782, 264288, 264768, 264769, 264632, 264555, 264639, 56526486, 22279000
1583	87622715 (3165, 3166)	Novel Protein sim. GBank gij5578958[emb CAB51351.1]- (AL050306) dJ475B7.2 (novel protein) [Homo sapiens]		UNCLASSIFIED	60170831, 33657402, 264682, 21906766, 35895855, 264563
1584	95337722 (3167, 3168)	Novel Protein sim. GBank gij5531815[gib AAD44482.1]- (AF078850) steroid dehydrogenase homolog [Homo sapiens]	Contains protein domain (PF00106) - short chain dehydrogenase	- dehydrogenase	60424179, 52646842, 65274572, 56182575, 22278995, 35696286, 22278996, 22278998, 22278999, 264259, 29331822, 56182181, 60424269, 60432289, 29331827, 29331828, 35896052, 29146498, 66712502, 29331830, 52644045, 56182435, 264510, 264512, 265008, 60433356, 33657402, 60433438, 55812038, 21906754, 55811386, 52644296, 87168474, 87168559, 265018, 265019, 264448, 264369, 264288, 18108358, 21906765, 21906767, 21906768, 21906769, 35895917, 265020, 265021, 265022, 52644150, 33657023, 33657109, 18108374, 55810764, 55811576, 35696423, 65274791, 35895855, 56182323, 83373044, 18108387, 87168518, 60432113, 22279002
1585	87626117 (3169, 3170)	Novel Protein sim. GBank gij4240132[dbj BAA74846.1]- (AB020630) KIAA0823 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	- phosphatase	35896286, 22278998, 264259, 29331822, 29331824, 29331825, 264905, 265006, 265007, 265008, 60433356, 33109954, 87168474, 265011, 265017, 264604, 264369, 264288, 264685, 264769, 18108359, 21906765, 18108384, 18108370, 264629, 263972, 18108383, 18108388, 264482, 264504
1586	88067081 (3171, 3172)	Novel Protein sim. GBank gij3786494 (AF098993) - No definition line found [Caenorhabditis elegans]			265017, 265018, 264689, 33657023, 263978, 264636, 264563
1587	87617126 (3173, 3174)	Novel Protein sim. GBank gij3253159 (AF005355) - translation initiation factor eIF2C [Oryctolagus cuniculus]		UNCLASSIFIED	264907, 264908, 264511, 264910, 264591, 264594, 264629, 264631, 264563, 264483, 264567
1588	87802536 (3175, 3176)	Novel Protein sim. GBank gij1077573[pir J52680 - probable ribosomal protein L34, mitochondrial - yeast (Saccharomyces cerevisiae)]	Contains protein domain (PF00468) - Ribosomal protein L34	UNCLASSIFIED	264259, 29331828, 264905, 265006, 264758, 21906754, 264761, 264762, 21906765, 21906769, 60170615, 52644150, 33657109, 35895855, 56182323, 18108385
1589	90980653 (3177, 3178)	Novel Protein sim. GBank gij2137756[pir J48746 - semaphorin C - mouse (fragment)]		UNCLASSIFIED	65274572, 264490, 29331822, 66714117, 29331827, 29331828, 56182435, 265008, 60170831, 264595, 264758, 264596, 265011, 264686, 21906766, 21906768, 55811957, 27466265, 264639, 18108385, 56526486, 60432113

1590	95319825 (3179, 3180)			UNCLASSIFIED	264489, 22278996, 264259, 29331824, 29331825, 29331826, 29331827, 265006, 60433356, 21906754, 265017, 265018, 265019, 264448, 264785, 264288, 52844229, 21906765, 21906767, 21906768, 21906769, 265021, 264692, 27486265, 35695763, 56526486, 60432113, 22278000, 22279002, 264554
1591	86877160 (3181, 3182)	Novel Protein sim. GBank gi4557749 ref NP_000237.1 pMHC2 - MHC class II transactivator		MHC	264259, 264905, 29331830, 264595, 265017, 264448, 264288, 264690, 264629, 87168518
1592	87882533 (3183, 3184)			UNCLASSIFIED	65274572, 60432049, 264509, 60433356, 21906754, 21906767, 21906768, 18108370, 35696423, 22279000, 264565, 264567
1593	94991661 (3185, 3186)			UNCLASSIFIED	264488, 29331827, 264905, 264906, 264907, 264908, 264909, 264910, 264592, 264593, 264757, 264602, 264604, 264760, 264681, 264288, 264766, 264768, 29148629
1594	87773752 (3187, 3188)	Novel Protein sim. GBank gi 3877072 emb CAA87060  - (Z48937) similarity with ribosomal protein L21 [Caenorhabditis elegans]		UNCLASSIFIED	35695917, 264692, 264628, 264629, 264630, 264632, 264634, 264635, 264638, 264639, 264563, 264584, 264586
1595	79819425 (3189, 3190)	Novel Protein sim. GBank gi 3152703 (AF065389) - tetraspan NET-4 [Homo sapiens]	Contains protein domain (PF00335) - 4 transmembrane segments integral membrane proteins	UNCLASSIFIED	29331826, 264908, 55811957
1596	78933928 (3191, 3192)			UNCLASSIFIED	29146498, 264758, 263967
1597	86971857 (3193, 3194)	Novel Protein sim. GBank gi 5257114 gb AAD41244.1 AF09448 - (AF094480) cholesterol 24-hydroxylase [Homo sapiens]	Contains protein domain (PF00067) - Cytochrome P450	Cyto450	264092, 29331824, 264508, 264682, 264389, 264686, 264630, 264563
1598	87862839 (3195, 3196)				264259, 264634
1599	87849829 (3197, 3198)	Novel Protein sim. GBank gi 4508797 ref NP_000324.1 pSCA7 - spinocerebellar ataxia 7 (olivopontocerebellar atrophy with retinal degeneration)	Contains protein domain (PF01581) - FMRFamide related peptide family	UNCLASSIFIED	52645080, 29331824, 29331826, 264511, 265009, 265011, 264605, 264448, 264764, 265020, 264692, 264693, 18108370, 264635, 18108385
1600	80056002 (3199, 3200)				29331826, 264603, 264691, 264563
1601	15023246 (3201, 3202)				264635
1602	86926987 (3203, 3204)	Novel Protein sim. GBank gi 5305704 gb AAD41780.1 AF12853 - (AF128535) cytoplasmic phosphoprotein PACSIN2 [Mus musculus]	Contains protein domain (PF00018) - SH3 domain	UNCLASSIFIED	29146499, 264112, 264762, 18108351, 29148627, 263974
1603	80502072 (3205, 3206)	Novel Protein sim. GBank gi 283920 pir S27939 - tensin - chicken		collagen	264490, 29331824, 264907, 264909, 264511, 265008, 264592, 265010, 265011, 264762, 264764, 264369, 264288, 264687, 264769, 264693, 264628, 264634, 264636, 264555, 264556, 264638, 264557, 264558, 264559, 18108385
1604	80221813 (3207, 3208)	Novel Protein sim. GBank gi 4768831 gb AAD29633.1 AF11682 - (AF116827) unknown [Homo sapiens]		ATPase-associated	18108385, 263977



1605	91221129 (3209, 3210)				struct	264905, 264509, 264906, 264907, 264908, 264909, 264604, 264766, 264768, 264692, 264693, 33657109, 264629, 35695855, 264635, 264636, 264637
1606	94312703 (3211, 3212)	Novel Protein sim. GBank gij4505313[ref]NP_003794.1[pMYOM - UNKNOWN	Contains protein domain (PF00047) - Immunoglobulin domain	struct	22278996, 22278999, 264259, 33657402, 265017, 18108351, 264448, 21906767, 21906769, 52644150, 264691, 87168518	
1607	10871805 (3213, 3214)	Novel Protein sim. GBank gij5174473[ref]NP_005888.1[pPPI] - Intracisternal A particle-promoted polypeptide		transcriptfactor	264689	
1608	80428900 (3215, 3216)	Novel Protein sim. GBank gij2224629[dbj]BAA20802] - (AB002342) KIAA0344 [Homo sapiens]		UNCLASSIFIED	264094, 264806, 264907, 264909, 264910, 264591, 264603, 264768, 264693, 264634, 264635, 264637, 264639	
1609	94311572 (3217, 3218)	Novel Protein sim. GBank gij4884073[emb]CAB43213.1] - (AL049934) hypothetical protein [Homo sapiens]			52644507, 52645156, 52646365, 52646842, 56182575, 22278994, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29147620, 29331828, 35696052, 33656970, 264508, 264509, 264907, 52644045, 56182435, 264510, 264511, 264512, 33657402, 21906754, 52646317, 33109954, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264448, 264288, 264769, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 52644150, 33657023, 33657109, 52645129, 27486261, 27486262, 35695763, 264628, 18108370, 18108376, 35696423, 264638, 52644332, 18108387, 87168518, 22278000, 264563, 264486	
1610	85468200 (3218, 3220)	Novel Protein sim. GBank gij283920[pir]IS27939 - tensin - chicken		UNCLASSIFIED	264593, 264757, 55812038, 265018, 265020, 264691, 264692, 264693, 264631, 264634, 264635, 264555, 22279000, 264564	
1611	94122843 (3221, 3222)	Novel Protein sim. GBank gij107284[pir]JA35415 - peroxidase (EC 1.11.1.7), thyroid (Grave's disease) - human (fragment)	Contains protein domain (PF00008) - EGF-like domain	peroxidase	35696286, 21908765, 264691, 35696423	
1612	85746031 (3223, 3224)	Novel Protein sim. GBank gij387484[emb]CAA94337] - (Z70307) Similarity to B.subtilis tetracycline resistance protein (SW:TCR2_BACSU); cDNA EST EMBL:C09951 comes from this gene; cDNA EST EMBL:C08265 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	264488, 264509, 18108370, 18108387, 264486	
1613	82247354 (3225, 3226)			UNCLASSIFIED	264759	

1614	91228634 (3227, 3228)	Novel Protein sim. GBank gi4680673 db AAD27726.1 AF132951 - (AF132951) CGI-17 protein [Homo sapiens]	Contains protein domain (PF01606) - eRF1-like proteins	UNCLASSIFIED	22278995, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 264908, 264512, 265009, 265011, 265017, 265018, 265019, 18108351, 264683, 264288, 264766, 21906767, 21906768, 21906769, 35695917, 265021, 265022, 35696423, 35695855, 60170394, 56182323, 83373044, 264566
1615	86121809 (3229, 3230)	Novel Protein sim. GBank gi15689485 db GAA83026.1  - (AB028997) KIAA1074 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	homeobox	22278996, 35696286, 22278997, 29331822, 35696052, 29331828, 264508, 264908, 264909, 56182435, 264511, 265017, 265019, 264766, 264767, 264768, 265020, 264691, 264628, 264632, 264635, 264555, 264556, 56182323, 264558, 22279002
1616	94311819 (3231, 3232)	Novel Protein sim. GBank gi3876260 emb CAB01696  - (Z78418) cDNA EST EMBL:D71020 comes from this gene; cDNA EST EMBL:D73593 comes from this gene; cDNA EST EMBL:C07649 comes from this gene; cDNA EST EMBL:C09081 comes from this gene; cDNA EST yk399f2.3 comes from this gene; cDNA ...	UNCLASSIFIED	UNCLASSIFIED	264488, 52644507, 52645158, 52646365, 52646842, 22278994, 22278995, 35696286, 22278996, 22278997, 22278999, 52645080, 29331822, 29331824, 29331825, 29331827, 29331828, 35696052, 35656970, 264905, 264909, 264594, 52646317, 21906754, 33657084, 52644296, 87168474, 87168559, 265017, 265018, 265019, 264681, 264448, 264684, 52644229, 21906764, 264689, 21906765, 21906766, 21906769, 35695917, 265020, 265021, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 33657349, 27486265, 35695763, 18108376, 35696423, 35695855, 264557, 52644332, 264558, 18108385, 87168518
1617	98090742 (3233, 3234)	Novel Protein sim. GBank gi466053 sp P34678 YO41_CAEEL - HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III	Contains protein domain (PF01529) - DHHC zinc finger domain	peptidase	35696052, 264905, 264509, 264907, 264908, 264510, 264511, 264764, 264766, 264768, 264689, 264693, 18108374, 264635, 264636, 264638
1618	86272860 (3235, 3236)	Novel Protein sim. GBank gi4240231 db GAA74894.1  - (AB020678) KIAA0871 protein [Homo sapiens]		struc	35696286, 22278999, 264092, 29331824, 29331825, 35696052, 33657084, 21906765, 27486264
1619	95354580 (3237, 3238)	Novel Protein sim. GBank gi15031763 ref NP_005515.1 pHRY  - hairy (Drosophila)- homolog	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	transcriptfactor	52646842, 65274572, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 56182435, 265007, 265008, 264910, 60170831, 60432229, 60433356, 60433438, 265019, 264448, 264288, 264686, 21906768, 265021, 60170615, 33657023, 65274620, 33657109, 18108374, 18108376, 35696423, 35695855, 56182323, 56526486
1620	87344655 (3239, 3240)	Novel Protein sim. GBank gi1351047 sp P45843 SCRT_DROME - SCARLET PROTEIN		UNCLASSIFIED	264684

1621	87076708 (3241, 3242)	Novel Protein sim. GBank gi 3822553 (AF098788) - nuclear calmodulin-binding protein [Gallus gallus]	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	264910 18108392, 65274572, 18108398, 22278996, 22278997, 22278999, 29146498, 29146499, 264905, 264908, 264909, 264928, 26444045, 264592, 60433356, 21906754, 264602, 265017, 264368, 21906768, 55811957, 265021, 60170615, 264635, 264557, 60170394, 83373044, 18108385, 22279000, 22278002, 264566
1622	94741739 (3243, 3244)	Novel Protein sim. GBank gi 731086 sp P40389 UV22_SCHPO - UV-INDUCED PROTEIN UV22		UNCLASSIFIED	18108398, 264259, 264909, 56182435, 87168474, 264448, 21906768, 35695917, 264691, 87168518, 264583
1623	87779106 (3245, 3246)	Novel Protein sim. GBank gi 3875666 emb CAB05478  - (Z83104) cDNA EST EMBL:TO0015 comes from this gene; cDNA EST EMBL:D33665 comes from this gene; cDNA EST EMBL:D36540 comes from this gene; cDNA EST yk2408.3 comes from this gene; cDNA EST yk387c8.3 comes from this gene; cDNA ES...		UNCLASSIFIED	66714117, 29331825, 264909, 265008, 264758
1624	87338178 (3247, 3248)	Novel Protein sim. GBank gi 4589622 db BAA76833.1  - (AB023206) KIAA0989 protein [Homo sapiens]		kinase	264488, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264259, 29331824, 29331825, 29331827, 29331828, 264102, 264108, 264508, 33657084, 265017, 265018, 18108351, 264683, 264369, 264288, 21906765, 21906766, 21906767, 21906769, 35695917, 265021, 264691, 65274620, 18108368, 263972, 18108376, 35696423, 264631, 264634, 22279000, 22279002
1625	94734369 (3251, 3252)	Novel Protein sim. GBank gi 5679070 gb AAD46844.1 JAF16090 - (AF160904) BcDNA.HL05936 [Drosophila melanogaster]			52644507, 52645156, 52646365, 52646842, 22278994, 56994075, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 66714117, 29331826, 60432289, 29331827, 29331828, 35696052, 264906, 66712502, 264909, 265008, 265009, 60432229, 60433356, 60433438, 21906754, 52646317, 52644296, 265011, 87168559, 264604, 265018, 264448, 264369, 264288, 264766, 52644229, 264689, 21906765, 21906768, 35695917, 265021, 265022, 52644150, 33657023, 65274620, 27486261, 27486262, 27486265, 35695763, 263972, 52644332, 60170394, 87168518, 60432113, 264557
1626	83368173 (3253, 3254)	Novel Protein sim. GBank gi 3668087 (AC004667) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264298
1628	85708459 (3255, 3256)			eph	264288, 264686, 264767, 22279002

1629	84993841 (3257, 3258)	Novel Protein sim. GBank gi 4240175 dbj BAA74866.1  - (AB020650) KIAA0843 protein [Homo sapiens]		struct	264555
1630	87779027 (3259, 3260)			UNCLASSIFIED	29331822, 29331827, 265010, 264693, 264634, 22279002
1631	87758454 (3261, 3262)	Novel Protein sim. GBank gi 1915892 emb CAA69995  - (Y08740) tom-1A protein [Gallus gallus]		UNCLASSIFIED	55811957, 264259, 33657023, 264693, 29331822, 29331824, 29331827, 29331828, 264906, 264908, 55811576, 264910, 264634, 264636, 264637, 56182323, 264559, 264758, 18108385, 264563, 264764, 264766
1632	87871692 (3263, 3264)	Novel Protein sim. GBank gi 2558501 dbj BAA22896  - (D63850) hepatoma-derived growth factor [Mus musculus]		UNCLASSIFIED	264687, 264769, 264691, 264692, 29146499, 284509, 264905, 264907, 284511, 284512, 284482, 264681, 264763, 264682, 264683
1633	87773683 (3265, 3266)				264488, 264259, 264907, 264908, 264909, 264628, 264629, 264631
1634	85992817 (3267, 3268)	Novel Protein sim. GBank gi 4887229 gb AAD32244.1 AF15075 - (AF150755) microtubule-actin crosslinking factor [Mus musculus]	Contains protein domain (PF00435) - Spectrin repeat	struct	265007, 264637, 22279002
1635	94232800 (3269, 3270)			UNCLASSIFIED	65274572, 22278998, 35696052, 52844045, 264511, 265008, 265009, 265010, 285011, 265018, 265019, 264448, 264369, 21908765, 21908768, 265021, 264690, 264482
1636	80413227 (3271, 3272)	Novel Protein sim. GBank gi 4557511 ref NP_001339.1 pDAPK - death-associated protein kinase 3	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	22278995, 264594, 264763, 265020
1637	80070435 (3273, 3274)				264558
1638	87101854 (3275, 3276)	Novel Protein sim. GBank gi 3420051 (AC004680) - unknown protein [Arabidopsis thaliana]			21906765, 21906767, 22278998, 35696286, 22278999, 264259, 264692, 264693, 29331824, 33657109, 264508, 264906, 18108370, 264629, 265007, 33657402, 21908754, 264602, 264604, 264764, 264683, 264566, 264288
1639	94322194 (3277, 3278)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264488, 18108394, 65274572, 56182575, 35696286, 29331824, 29331826, 29331827, 35696052, 264908, 56182435, 264112, 265006, 265008, 264757, 264758, 55811386, 264603, 264760, 18108351, 264764, 264288, 264766, 264768, 21906767, 55811957, 264691, 33657023, 65274620, 18108370, 55810764, 55811576, 264558, 264639, 83373044, 18108385, 87168518

1640	94143165 (3279, 3280)	Novel Protein sim. GBank gi 2842489 emb CAA16847.1  - (AL021747) hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	56182575, 56994075, 35696286, 60432049, 60432288, 29331827, 35698052, 52644045, 56182435, 264510, 265006, 265007, 265008, 264910, 265009, 33657402, 55812038, 265010, 265011, 265017, 265018, 264288, 52644229, 21906765, 21906766, 21906768, 35695917, 265021, 60170615, 52644150, 33657023, 33657109, 33657349, 18108374, 35696423, 65274791, 35695855, 264632, 264555, 56182323, 22279000
1641	87625160 (3281, 3282)			UNCLASSIFIED	29146499, 265006, 265007, 265008, 265009, 55812038, 265010, 265011, 264555, 264556, 264558, 18108383
1642	94312557 (3283, 3284)	Novel Protein sim. GBank gi 1575333 (U60416) - myr 6 myosin heavy chain [Rattus norvegicus]	Contains protein domain (PF01843) - struct DIL domain		22278989, 29147620, 29331826, 29331828, 33656970, 55812038, 265010, 265018, 265019, 18108351, 264689, 265020, 265022, 264690, 33657023, 65274620, 35695763, 52644332, 18108381, 60170394, 56182323, 18108388, 87168518, 22279002, 264564
1643	94131766 (3285, 3286)				29331825, 29331827, 29331828, 21906754, 265019, 264288, 264693, 33657349, 18108370, 18108376, 264555, 83373044, 22278002, 264482
1644	88095125 (3287, 3288)			UNCLASSIFIED	264905, 264907, 264908, 264910, 265009, 264757, 264758, 264761, 264762, 264763, 264766, 264768, 264769, 264628, 264629, 264630, 264631, 264632, 264563, 264564, 264565, 264566, 264567
1645	95013858 (3289, 3290)			UNCLASSIFIED	264685, 264693
1646	95362691 (3291, 3292)	Novel Protein sim. GBank gi 1076802 pir S49915 - extensin like protein - maize		UNCLASSIFIED	22278994, 56994075, 35696286, 264259, 29331824, 29331825, 29331826, 60432289, 264508, 60433356, 60433438, 87168559, 265018, 264687, 35695917, 264692, 33657023, 33657182, 27486261, 27486265, 33657349, 60432113, 264563, 264564
1647	94278428 (3293, 3294)	Novel Protein sim. GBank gi 5002573 emb CAB44338.1  - (Y17466) alpha-N-acetylgalactosamine alpha-2,6-sialyltransferase [Fugu rubripes]		UNCLASSIFIED	29331822, 264906, 264908, 264369, 21906768, 60170615, 264639, 22279000
1648	87642098 (3295, 3296)		Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	265009, 264686, 55811957, 35695917, 55810764, 264556, 56182323, 264558, 18108385

1649	95347628 (3297, 3298)	Novel Protein sim. GBank gijl854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 6]		cadherin	264488, 22278995, 35696286, 22278996, 22278997, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264907, 66712502, 264908, 52644045, 264909, 56182435, 264511, 265007, 265008, 265009, 264591, 264593, 60433438, 264596, 55812038, 21906754, 265011, 264601, 264602, 265017, 265018, 265019, 264682, 264448, 264764, 264683, 264288, 264768, 264685, 264687, 264768, 264688, 264769, 52644229, 264689, 21906765, 21906766, 21906767, 21906768, 55811957, 35695917, 265021, 265022, 52644150, 264692, 33657023, 33657109, 20281149, 18108370, 264628, 18108374, 18108376, 35696423, 35695855, 264632, 264634, 264635, 264636, 18108380, 264639, 264558, 18108382, 18108384, 18108385, 18108387, 264080, 264404, 60432113, 22279000, 22279002, 264482, 264565, 264566, 264487, 265011, 264602, 21906767, 18108374, 18108377, 18108385
1650	87418539 (3299, 3300)	Novel Protein sim. GBank gijl3647335[emb]CAA21059] - (AL031644) possible zinc-finger protein [Schizosaccharomyces pombe]			
1651	91639773 (3301, 3302)	Novel Protein sim. GBank gijl4884278[emb]CAB43247.1] - (AL050037) hypothetical protein [Homo sapiens]		synthase	264488, 52645156, 18108397, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264508, 264808, 29331830, 264910, 60432229, 21906754, 265010, 265011, 265017, 265019, 264448, 18108354, 264288, 264888, 21906765, 21906766, 21906768, 21906769, 265022, 264692, 264693, 264629, 35695855, 264556, 264637, 264557, 264559, 83373044, 56526486, 22279000, 22279002, 264564
1652	86598622 (3303, 3304)	Novel Protein sim. GBank gijl1657837 (U73200) - p116Rip [Mus musculus]		struct	22278997, 29146498, 56182435, 21906754, 264369, 21906765, 21906768, 21906769, 265020, 52644150, 33657109, 22279000, 22279002
1653	94255993 (3305, 3306)	Novel Protein sim. GBank gijl3776054[emb]CAA06273] - (AJ004999) Tapasin [Gallus gallus]		glycoprotein	18108398, 22278995, 22278998, 264259, 29331822, 29331824, 60432289, 29331826, 29331827, 29331830, 264909, 265006, 265009, 60432229, 60433356, 60433438, 21906754, 265017, 265019, 264448, 264683, 264288, 265021, 265022, 264692, 18108364, 65274791, 18108384, 60432113, 264567
1654	79756471 (3307, 3308)			UNCLASSIFIED	33657109, 264585

1655	86689346 (3308, 3310)	Novel Protein sim. GBank gi 3355717 emb CAA73496  - (Y13053) seryl-tRNA synthetase [Zea mays]		synthase	52644507, 35696286, 22278998, 22278999, 29331824, 29331825, 29331828, 33656970, 264908, 52644045, 264511, 264910, 52646317, 264288, 52644229, 33657023, 33657109, 52644332, 264557, 56162323, 56526486, 60432113
1656	79962297 (3311, 3312)	Novel Protein sim. GBank gi 1890141 dbj BAA18947  - (D83206) P24 protein [Mus musculus]		UNCLASSIFIED	29331822, 29331824, 29331825, 264563
1657	87771994 (3313, 3314)	Novel Protein sim. GBank gi 4557645 ref NP_001524.1 p-INRP - heterologous nuclear ribonucleoprotein L	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	29331827, 265009, 21906766, 21906767, 265020, 265022, 33657109, 264638, 56526486, 264482
1658	87773778 (3315, 3316)	Novel Protein sim. GBank gi 3877072 emb CAA87060  - (Z46937) similarity with ribosomal protein L21 [Caenorhabditis elegans]	Contains protein domain (PF00829) - Ribosomal protkaryotic L21 protein	UNCLASSIFIED	52646365, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 29146498, 264905, 264908, 52644045, 265005, 60433356, 264757, 60433438, 21906754, 265011, 18108351, 264448, 264369, 264288, 264766, 264768, 21906765, 21906767, 21906768, 21906769, 29148629, 265021, 265022, 18108362, 263968, 263971, 18108374, 35696423, 18108383, 22279000, 264482
1659	88230101 (3317, 3318)	Novel Protein sim. GBank gi 539218 prj J38038 - hypothetical protein YKL201c - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED	52646317, 21906766, 21906767, 21906768, 87168518, 22278996, 265020, 22278999, 87168559, 264603, 265017, 264631, 265018, 265019, 22279002, 264482, 264635, 264585
1660	94315313 (3319, 3320)	Novel Protein sim. GBank gi 2497012 sp Q10010 YSV4_CAEEL - HYPOTHETICAL 26.6 KD PROTEIN T19C3.4 IN CHROMOSOME III		UNCLASSIFIED	264488, 35696286, 264259, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265009, 264910, 264592, 264596, 265010, 264600, 264602, 265017, 265018, 264605, 264760, 264764, 264288, 264766, 264886, 264768, 264769, 264689, 21906766, 35695917, 264690, 33657023, 264693, 33657109, 264629, 35696423, 35695855, 264634, 264635, 264555, 264636, 264637, 264556, 264638, 264639, 264559, 18108385, 18108388, 264563, 264483, 264564, 264565, 264566, 264486, 264567

1661	94234071 (3321, 3322)	Novel Protein sim. GBank gi14759100 refNP_004759.1 pSFRS - splicing factor, arginine/serine-rich 11	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264488, 22278998, 264259, 29331824, 29331826, 29331827, 29331828, 264509, 66712502, 29331830, 264908, 52644045, 265007, 264512, 60433356, 60433438, 55812038, 21906754, 265019, 264448, 264766, 264768, 264768, 21906768, 21906769, 265020, 33657023, 33657109, 65274791, 87168518, 264482, 264563, 264564, 264565, 264567
1662	94135172 (3323, 3324)	Novel Protein sim. GBank gi11730502 sp PF52875 PF27_MOUSE - TRANSMEMBRANE PROTEIN PFT27			18108392, 29331822, 29331828, 20281100, 264106, 265008, 265007, 265008, 18108348, 21906766, 18108365, 18108366, 18108374, 83373044, 18108385
1663	94217146 (3325, 3326)	Novel Protein sim. GBank gi14884136 emb CAB43275.1  - (AL050107) hypothetical protein [Homo sapiens]	Contains protein domain (PF00397) - kinase WW domain	kinase	52645156, 56182575, 22278994, 22278995, 35696286, 22278996, 56994075, 22278997, 22278998, 22278999, 264259, 29331822, 29331826, 29331827, 29331828, 33656970, 29331830, 264908, 56182435, 264511, 60433356, 33657402, 33109954, 87168474, 87168559, 265017, 265018, 264605, 18108351, 264764, 264288, 264768, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 265022, 264691, 33657023, 264693, 263967, 33657109, 264630, 52644332, 83373044, 87168518, 60432113, 22278900
1664	94234078 (3327, 3328)	Novel Protein sim. GBank gi13043892 dbj BAA25510  - (AB011156) KIAA0584 protein [Homo sapiens]		UNCLASSIFIED	264488, 263994, 35696286, 29331824, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 60170831, 264591, 264592, 264595, 87168474, 265011, 264600, 264601, 264604, 264605, 264760, 264762, 18108351, 264681, 264682, 264763, 264683, 264764, 264288, 264684, 264766, 264687, 264768, 264769, 21906764, 21906765, 21906767, 35695917, 265021, 264534, 60170815, 264690, 264691, 264692, 33657109, 33657182, 264628, 18108370, 264629, 35696423, 35695855, 264634, 264635, 264555, 264638, 264637, 264638, 264639, 264558, 83373044, 87168518, 264563, 264566, 264486



1665	91228652 (3329, 3330)	Novel Protein sim. GBank gij1083505[pil][S50065 - slabadhesin - mouse	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	264488, 29331826, 29331828, 264509, 264906, 264907, 264909, 264510, 264511, 264910, 264592, 264593, 264595, 264758, 264596, 264600, 264760, 264762, 264764, 264766, 264768, 264628, 264630, 264634, 264636, 83373044, 264564, 264566, 264567, 264486
1666	95358160 (3331, 3332)	Novel Protein sim. GBank gij3913431[sp]Q42643[DDX8 SCHPO - PUTATIVE PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE SPAC10F6.02C	Contains protein domain (PF00575) - S1 RNA binding domain	helicase	56994075, 22278989, 264259, 29331824, 29331826, 29331827, 29146498, 265009, 33109954, 87168559, 265019, 264288, 264688, 21906767, 21906769, 264691, 33657182, 18108370, 18108374, 18108385, 22279002
1667	91228655 (3333, 3334)	Novel Protein sim. GBank gij5689535[dbj][BAA83051.1] - (AB029022) KIAA1089 protein [Homo sapiens]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	transport	264259, 29331822, 29331826, 264805, 264906, 264908, 264510, 265009, 264595, 264758, 265011, 87168559, 265017, 265018, 265019, 264448, 264768, 264686, 21906765, 21906767, 21906769, 265020, 265021, 60170615, 264690, 264692, 264693, 18108368, 18108370, 263972, 55810764, 264555, 83373044, 60432113, 22279000, 22279002
1668	88095135 (3335, 3336)	Novel Protein sim. GBank gij20768894[gbj][AAB53983.1] - (AF002197) short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG_PE-bind). Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]		kinase	66714117, 264508, 264509, 264906, 264907, 264908, 264511, 264910, 264764, 264687, 264689, 33657109, 35698423, 35695855, 264632
1669	91227848 (3337, 3338)	Novel Protein sim. GBank gij3875371[emb]CAA85414.1] - (Z36948) contains a valine and arginine rich domain, possesses weak similarity with the RNA binding domains from RNA splicing factor U2AF 65 KD subunit: cDNA EST EMBL:D64658 comes from this gene; cDNA EST EMBL:D66829 comes from this gene		UNCLASSIFIED	29331825, 33109954, 264369, 264767, 264689, 33657109, 83373044
1670	87628009 (3339, 3340)			UNCLASSIFIED	264259, 29331824, 29331827, 60433438, 265022, 264638
1671	87346372 (3341, 3342)	Novel Protein sim. GBank gij482451[sp]P34244[KKK1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	35696286, 22278997, 29331825, 264909, 21906754, 265017, 265018, 265019, 264682, 264683, 264768, 264688, 21906768, 21906767, 21906769, 264691, 264555, 264556, 22279000, 264566
1672	86291834 (3343, 3344)	Novel Protein sim. GBank gij1814270 (U74586) - double-stranded RNA specific adenosine deaminase [Rattus norvegicus]		deaminase	264906, 264909, 264632, 18108381

1673	88095137 (3345, 3348)	Novel Protein sim. GBank gi 2076894 gb AAB53983.1  - (AF002197) short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG_PE_bind), Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00130) - Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	264488, 264569, 18108394, 56994075, 22278996, 264259, 35696052, 264508, 264905, 264509, 264906, 264507, 264908, 264909, 264510, 264910, 60170831, 264592, 264594, 264595, 264758, 264601, 264760, 264762, 264683, 264764, 264288, 264766, 264686, 264768, 264687, 264769, 264689, 264690, 33657023, 264692, 264693, 33657109, 264628, 264629, 18108374, 35696423, 35695855, 264631, 264632, 264634, 264635, 264637, 264556, 264638, 264639, 264563, 264482, 264564, 264565, 264566, 264567, 264486
1674	88258028 (3347, 3348)	Novel Protein sim. GBank gi 5262467 emb CAB45693.1  - (AL080082) hypothetical protein [Homo sapiens]		kinase	29331822, 29331824, 264906, 52644045, 60433356, 87168559, 264448, 264288, 264686, 264691
1675	87606466 (3349, 3350)	Novel Protein sim. GBank gi 3128366 (AF010496) - 50S ribosomal protein l9 [Rhodospirillum rubrum]		UNCLASSIFIED	56181686, 35696286, 22278997, 22278998, 264259, 29331824, 29331827, 35696052, 66712502, 264764, 264288, 264686, 264687, 35695917, 265020, 264690, 264693, 35695763, 18108370, 35696423, 35695855, 264637, 264639, 18108385, 264564
1676	95358086 (3351, 3352)	Novel Protein sim. GBank gi 4164065 gb AAD05327  - (AF111091) latrophilin 3 splice variant bbat [Bos taurus]		UNCLASSIFIED	264259, 29331827, 29331828, 264106, 264907, 265009, 264600, 265019, 264288, 21906765, 265020, 265022, 35695855, 83373044, 18108385
1677	87408587 (3353, 3354)	Novel Protein sim. GBank gi 3327046 dbj BAA31591  - (AB014516) KIAA0616 protein [Homo sapiens]		UNCLASSIFIED	264908
1678	86866829 (3355, 3356)	Novel Protein sim. GBank gi 550452 (U08469) - 3-methylcrotonyl-CoA carboxylase, biotin-carrier domain [Glycine max]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSase)	UNCLASSIFIED	29331824, 264102
1679	91214106 (3357, 3358)			carboxylase	264488, 18108392, 18108394, 52646842, 18108397, 18108398, 35696286, 29331824, 265006, 265007, 265008, 265009, 18108348, 265011, 18108351, 264683, 18108354, 18108358, 18108359, 21906765, 29148627, 29148629, 264690, 18108361, 18108362, 18108364, 18108365, 18108368, 264628, 18108379, 35696423, 35695855, 264635, 18108381, 18108382, 18108383, 18108384, 18108385, 18108388
1680	91005372 (3359, 3360)	Novel Protein sim. GBank gi 2394478 (AF024500) - No definition line found [Caenorhabditis elegans]		transport	65274572, 22278994, 22278999, 66714117, 29331827, 56182435, 21906754, 265018, 264288, 21906769
1681	94324150 (3361, 3362)	Novel Protein sim. GBank gi 5689537 dbj BAA83052.1  - (AB029023) KIAA1100 protein [Homo sapiens]		UNCLASSIFIED	22278996, 29331822, 264908, 264593, 264604, 265019, 264683, 55811957, 264690, 33657023, 35696423, 83373044, 264563
1682	86042710 (3363, 3364)			UNCLASSIFIED	264909, 265017, 264605

1683	94316213 (3365, 3366)	Novel Protein sim. GBank gij5031717ref[NP_005704.1]pGPBP - goodpasture antigen- binding protein	Contains protein domain (PF01852) - START domain	UNCLASSIFIED	263994, 35696286, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265006, 265007, 265009, 264910, 264593, 264758, 265010, 285011, 264760, 264761, 264762, 264763, 264764, 264369, 264766, 264768, 35695917, 264692, 33657109, 35696423, 35695855, 264634, 264635, 264636, 264638, 264639, 83373044, 264486
1684	80063409 (3367, 3368)			UNCLASSIFIED	264563, 264566
1685	94323182 (3369, 3370)	Novel Protein sim. GBank gij1255371 (U53147) - coded for by C. elegans cDNA yk34a9.5; coded for by C. elegans cDNA yk34a9.3; Similar to guanylate kinase. [Caenorhabditis elegans]	Contains protein domain (PF00825) - Guanylate kinase	UNCLASSIFIED	60424179, 52846842, 22278994, 35699286, 22278998, 264259, 52845080, 29331824, 29331826, 265007, 33657084, 265018, 264681, 264448, 264683, 264369, 264689, 21906765, 21906767, 21906768, 21906769, 265021, 264692, 65274620, 33657109, 27486262, 264635, 52644332, 56182323, 22279000
1686	87820710 (3371, 3372)	Novel Protein sim. GBank gij2244707[dbj]BAA21115.1] - (AB005287) thrombospondin 1 [Bos taurus]		UNCLASSIFIED	56182575, 264259, 264508, 264905, 264509, 264907, 264908, 264510, 264511, 265006, 264512, 265009, 264910, 264758, 265010, 265011, 264805, 18108351, 264784, 264766, 18108357, 264768, 18108362, 264628, 264630, 264631, 264634, 264635, 264637, 264638, 264639, 264565, 264486, 264567
1687	94719400 (3373, 3374)	Novel Protein sim. GBank gij4680679[gbj]AAD27729.1[AF132954] CGI-20 protein [Homo sapiens]		UNCLASSIFIED	35695917, 264906, 264907, 264908, 264510, 265006, 265007, 264910, 264556, 18108381, 18108383, 265011
1688	82158442 (3375, 3376)			UNCLASSIFIED	264905, 264910, 264760, 264629, 264555
1689	94325049 (3377, 3378)	Novel Protein sim. GBank gij4240193[dbj]BAA74875.1] - (AB020659) KIAA0852 protein [Homo sapiens]		UNCLASSIFIED	264569, 35696286, 22278999, 264508, 264908, 264909, 56182435, 33657402, 60433438, 55812038, 285017, 265018, 264448, 264764, 264288, 264686, 29148629, 35695917, 265020, 265021, 263972, 18108374, 65274791, 83373044, 264089
1690	83255346 (3379, 3380)	Novel Protein sim. GBank gij3800736 (AF031572) - seven- pass transmembrane receptor precursor [Mus musculus]	Contains protein domain (PF00028) - Cadherin domain	cadherin	264369, 21906766, 264692, 264639, 87168518
1691	88095223 (3381, 3382)	Novel Protein sim. GBank gij2773208 (AF039713) - No definition line found [Caenorhabditis elegans]			264768, 33657109, 29331827, 29148629, 264510, 264106, 264910, 264109, 264508, 50170831, 264563, 264905, 264564, 264691, 264637, 264628, 264907, 264908, 33657023, 264567, 264768, 263974
1692	86106709 (3383, 3384)				264106
1693	87012775 (3385, 3386)	Novel Protein sim. GBank gij121271[sp]P02207[GLB_LAMFL - GLOBIN	Contains protein domain (PF00042) - Globin	UNCLASSIFIED	29331826, 264508, 264905, 264907, 264595, 265010, 265011, 21906768, 33657023, 264629, 263978, 264558

1694	94208168 (3387, 3388)	Novel Protein sim. GBank gi 5453932 refNP_006225.1 PPOLR - polymerase (RNA) II (DNA directed) polypeptide J (13.3kD)	Contains protein domain (PF01193) - RNA polymerases L / 13 to 16 kDa subunit	35696286, 22278998, 22278998, 22278998, 264259, 29331822, 29331825, 29331826, 29331828, 35696052, 29146499, 264905, 264908, 52644045, 264511, 285008, 265007, 265009, 264592, 60433356, 21906754, 265010, 265011, 18108351, 264763, 264682, 264448, 264683, 264288, 264768, 264689, 21906768, 60170615, 264691, 264692, 264693, 18108370, 18108374, 263978, 35696423, 35695855, 264558, 18108381, 18108385, 87168518, 264482, 264486 264634
1695	94719325 (3388, 3390)	Novel Protein sim. GBank gi 4680679 gb AAD27729.1 AF13295 - (AF132954) CGI-20 protein [Homo sapiens]	UNCLASSIFIED	
1696	87824038 (3391, 3392)	Novel Protein sim. GBank gi 4220517 emb CAA22990  - (AL035356) hypothetical protein [Arabidopsis thaliana]		22278997, 264258, 265010, 18108351, 264764, 21906756, 18108370 264682
1697	85740963 (3393, 3394)	Novel Protein sim. GBank gi 505652 (U10362) - GP36b glycoprotein [Homo sapiens]	glycoprotein	
1698	87445285 (3395, 3396)	Novel Protein sim. GBank gi 5052031 gb AAD38411.1 AF15573 - (AF155739) axotrophin [Mus musculus]		56994075, 22278998, 22278999, 264509, 33657402, 264758, 87168474, 87168559, 265017, 265018, 264448, 264687, 29148627, 21906769, 29148629, 265020, 265022, 33657023, 264558, 87168518, 22279002 264635
1699	87424793 (3397, 3398)	Novel Protein sim. GBank gi 543344 pir S41647 - zinc finger 5 protein - mouse	UNCLASSIFIED nucl_rept	29331824, 52644045, 265008, 265009, 263969, 263971 264092, 264110, 263977
1700	87859161 (3399, 3400)	Novel Protein sim. GBank gi 3877439 emb CAA96652  - (Z72510) similarity to yeast UTR3 protein (Swiss Prot accession number P21374); cDNA EST EMBL:D72822 comes from this gene; cDNA EST EMBL:D75763 comes from this gene; cDNA EST yk274e3.3 comes from this gene; cDNA EST yk274e3.5 c...	MHC	22278995, 22278997, 264092, 29146498, 29146499, 264107, 264508, 264907, 264110, 264112, 265009, 60170831, 21906754, 265011, 265017, 264762, 18108351, 264288, 21906765, 35695917, 265021, 60170815, 263967, 33657109, 18108370, 263972, 263974, 18108374, 263976, 35695855, 264555, 263981, 60170384, 18108385, 56526486, 87168518, 60432113 264909, 265017, 264628, 264629, 264638
1701	86570488 (3401, 3402)	Novel Protein sim. GBank gi 451544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	UNCLASSIFIED	
1702	87795092 (3403, 3404)	Novel Protein sim. GBank gi 1263289 (U47856) - fibroin-4 [Araneus diadematus]	UNCLASSIFIED	264369
1703	79588651 (3405, 3406)	Novel Protein sim. GBank gi 4519621 dbj BAJ75670.1  - (AB017614) OASIS protein [Mus musculus]		264569, 35696286, 264907, 265010, 264687, 264768, 264692, 264693, 264636, 264568 22278996, 22278998, 264259, 264509, 265018, 264764, 264685, 264686, 21906768, 21906769, 265022, 264691, 264558, 22279000
1704	86622979 (3407, 3408)	Novel Protein sim. GBank gi 1263289 (U47856) - fibroin-4 [Araneus diadematus]		
1705	87795175 (3409, 3410)	Novel Protein sim. GBank gi 4519621 dbj BAJ75670.1  - (AB017614) OASIS protein [Mus musculus]		
1706	87790967 (3411, 3412)	Novel Protein sim. GBank gi 3123034 sp Q15011 Y025_HUMAN - HYPOTHETICAL PROTEIN KIAA0025		

1707	88041230 (3413, 3414)	Novel Protein sim. GBank gi4321664[gb AA015797] - (AF055470) ZNF258 [Homo sapiens]	UNCLASSIFIED	18108396, 22278997, 264259, 29147620, 29331826, 29146498, 264905, 264906, 265008, 264593, 264595, 264758, 264596, 265018, 264760, 18108351, 264764, 264766, 264689, 264693, 18108370, 35698423, 55811576, 264558, 87168518, 60432113, 264567
1708	91220519 (3415, 3416)	Novel Protein sim. GBank gi5174591[ref NP_005947.1 pMTfH - 5,10-methylene-tetrahydrofolate dehydrogenase, 5,10-methylene-tetrahydrofolate cyclohydrolase, 10-formyl-tetrahydrofolate synthetase	Synthase	56182575, 22278996, 56994075, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 29146498, 29146499, 29331830, 265009, 60170831, 33657402, 33109854, 87168559, 265019, 18108351, 264448, 21906765, 21906767, 21906768, 29148627, 29148629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108385, 22279000, 264563, 264107, 55811957, 263974, 263976, 263977, 263981
1709	80222563 (3417, 3418)		UNCLASSIFIED	264556
1710	20754572 (3419, 3420)	Novel Protein sim. GBank gi5031735[ref NP_005760.1 pHEC - N-acetylglucosamine 6-O-sulfotransferase	UNCLASSIFIED	65274572, 29331824, 29331826, 264768, 60431528, 35698423, 80432113, 264563
1712	95330184 (3423, 3424)	Novel Protein sim. GBank gi5454168[ref NP_006453.1 pXAP4 - HBV associated factor	kinase	56994075, 264093, 264259, 29331822, 264099, 29331824, 29331827, 264107, 264110, 264511, 264592, 265011, 265018, 264683, 264686, 264689, 265020, 33657023, 263967, 33657109, 263974, 35698423, 35695855, 264630, 264636, 264558, 264566
1713	94143453 (3425, 3426)	Novel Protein sim. GBank gi160409 (M59183) - mature-parasite-infected erythrocyte surface antigen [Plasmodium falciparum]	UNCLASSIFIED	22278995, 264508, 264758, 18108351, 18108370, 263974, 18108374, 264634, 56182323, 83373044, 60432113
1714	87420048 (3427, 3428)			22278997, 264757, 21906765, 265020, 265021, 264692, 56526486
1715	94260257 (3429, 3430)	Novel Protein sim. GBank gi5889537[dbj BAA83052.1  - (AB029023) KIAA1100 protein [Homo sapiens]	UNCLASSIFIED	264509, 264905, 264906, 264907, 264908, 264909, 264910, 264591, 265011, 264766, 264768, 264769, 264691, 264692, 264632, 264634, 264635, 264636, 264637, 264556, 264639, 264564
1716	87400449 (3431, 3432)	Novel Protein sim. GBank gi4589468[dbj BAA76761.1  - (AB012808) mBOCT [Mus musculus]	transport	56182575, 29331824, 60432289, 264109, 264909, 265007, 264600, 265019, 264686, 265020, 264693, 55811576, 264558, 60432113, 22279002

1717	87563223 (3433, 3434)	Novel Protein sim. GBank gi 2765411 emb CAA74749  - (Y14391) GTP-binding protein [Homo sapiens]		UNCLASSIFIED	264569, 264259, 29331825, 29331826, 29331828, 35696052, 264509, 264905, 264907, 264908, 264909, 264512, 265009, 264910, 264592, 264758, 264759, 265017, 264681, 264764, 264766, 264686, 18108357, 35695917, 264690, 264692, 264693, 264628, 264629, 35696423, 264630, 264631, 264635, 264636, 18108380, 264638, 264639, 18108388, 18108391
1718	87032628 (3435, 3436)	Novel Protein sim. GBank gi 2833262 sp Q14999 Y076_HUMAN - HYPOTHETICAL PROTEIN KIAA0076 (HA0936)		UNCLASSIFIED	265011, 264681, 264682, 264684, 264688, 264689, 21906765, 265021, 264691, 33657023, 264693, 18108370, 35695855, 264632, 264634, 264636, 18108388, 22278002
1719	94315259 (3437, 3438)	Novel Protein sim. GBank gi 4505197 ref NP_003473.1 pMML2 - myeloid/lymphoid or mixed-lineage leukemia 2		UNCLASSIFIED	18108396, 65274572, 35696286, 22278997, 60432049, 56182181, 66714117, 60432289, 29331828, 35696052, 29331828, 264906, 29331830, 56182435, 264592, 60431735, 60433438, 55812038, 264759, 265010, 264600, 264601, 265017, 264448, 264764, 264288, 264769, 21906766, 21906769, 55811957, 265020, 265021, 52644150, 33657023, 33657109, 33657182, 27486262, 33657349, 35695763, 18108370, 60431528, 18108374, 35696423, 55811576, 35695855, 264631, 56182323, 264559, 264564, 264486, 56182575, 22278999, 264259, 29331824, 60432289, 29331827, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 264758, 21906754, 265011, 264601, 264760, 264762, 264288, 264768, 264886, 18108357, 264689, 21906765, 55811957, 264693, 20281149, 264629, 18108374, 55811576, 65274791, 264630, 20281071, 264634, 264635, 264636, 264637, 264556, 264638, 264639, 56182323, 87168518
1720	94853063 (3439, 3440)	Novel Protein sim. GBank gi 2129478 pir S51839 - chitinase (EC 3.2.1.14) precursor - beet		UNCLASSIFIED	22278994, 22278999, 29331822, 265006, 265007, 265008, 55812038, 21906754, 60174639, 265011, 87168559, 18108351, 18108354, 21906765, 21906766, 21906768, 21906769, 265020, 33657109, 18108370, 18108374, 264556, 60170394, 83373044, 18108385, 264486
1721	91722288 (3441, 3442)	Novel Protein sim. GBank gi 4886461 emb CAB43381.1  - (AL050280) hypothetical protein [Homo sapiens]		UNCLASSIFIED	36994075, 29331824, 29331828, 265009, 18108351, 21906768, 265020, 33657023, 18108374, 83373044
1722	94134549 (3443, 3444)	Novel Protein sim. GBank gi 5689375 dbj BAA82958.1  - (AB030644) tudor repeat associator with PCTAIRE 2 [Rattus norvegicus]	Contains protein domain (PF00567) - kinase Tudor domain		

1723	95356181 (3445, 3446)	Novel Protein sim. GBank gi 426962 gb AAD20633  - (AF126062) Arf-like 2 binding protein BART1 [Homo sapiens]		UNCLASSIFIED	264488, 264687, 264769, 21906767, 21906768, 56182575, 55811957, 22278997, 22278998, 265020, 264259, 264692, 33657023, 29331822, 29331824, 29331825, 60432289, 33657182, 33856970, 33857349, 29146499, 264508, 264907, 18108370, 264629, 264908, 264909, 18108374, 55811576, 264510, 265006, 264511, 265007, 264910, 264632, 264591, 60432229, 264592, 60433356, 264594, 60433438, 264595, 83373044, 55812038, 33109954, 33657084, 87168518, 87168474, 285010, 285011, 87168559, 264600, 60432113, 264604, 285019, 264563, 264448, 264682, 264586, 264764, 264288, 264567, 264488, 264389, 264786
1724	87713808 (3447, 3448)	Novel Protein sim. GBank gi 2340162 (AF005083) - dsRBP-ZFa [Xenopus laevis]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	264905, 18108359, 264693, 264628, 264631, 264638, 264555, 264558, 264558, 264559
1725	85655191 (3449, 3450)	Novel Protein sim. GBank gi 3152862 (AF064604) - KE03 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	homeobox	35696286, 264259, 29331822, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264910, 265009, 264591, 264601, 264760, 18108351, 264681, 264764, 264288, 264768, 264788, 21906769, 35695917, 264628, 35696423, 264630, 264631, 264632, 264635, 264638, 264638, 87168518, 264566
1726	85754255 (3451, 3452)	Novel Protein sim. GBank gi 4689348 gb AAD27861.1 AF132562 (AF132562)		UNCLASSIFIED	29146498, 264683, 264689
1727	85296362 (3453, 3454)	BcDNA, LD14270 [Drosophila melanogaster]		UNCLASSIFIED	264905, 265011, 264689, 21906768

1728	95349515 (3455, 3456)	Novel Protein sim. GBank gij4406549[gb AAD20027] - (AF131738) Unknown [Homo sapiens]		UNCLASSIFIED	60424179, 18108397, 56182575, 22278995, 56994075, 35696286, 22278997, 22278998, 22278999, 264094, 60432049, 264259, 29331822, 29331824, 56182181, 29331825, 60432289, 29331826, 29331827, 35696052, 264905, 264906, 264907, 29331830, 66712502, 264908, 56182435, 264511, 265008, 265009, 60432229, 60433356, 33657402, 60433436, 264759, 21908754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 55811150, 264681, 264448, 264682, 264763, 264683, 264288, 264684, 264369, 264685, 264766, 264687, 264769, 21906764, 264689, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 265021, 265022, 264535, 264681, 264692, 33657023, 264693, 33657109, 18108370, 264628, 263972, 264629, 18108374, 18108376, 55810764, 65274791, 35695855, 264631, 264634, 264635, 60431850, 264636, 264638, 60170394, 264639, 83373044, 56526486, 87168518, 60432113, 22279000, 22279002, 264564, 264566
1729	91227948 (3457, 3458)	Novel Protein sim. GBank gij854065[emb CAA58337] - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264906, 264907, 264908, 264511, 264555, 83373044, 264596, 264566
1730	85463474 (3459, 3460)			UNCLASSIFIED	29331822, 29331825, 29331828, 264907, 264908, 264909, 265011, 264764, 264629
1731	88266068 (3461, 3462)	Novel Protein sim. GBank gij631600[pir S47094 - hypothetical protein - rabbit]		UNCLASSIFIED	52646842, 264907, 264909, 56182435, 55811386, 87168559, 265018, 265019, 264760, 52644229, 55811576
1732	91218878 (3463, 3464)	Novel Protein sim. GBank gij4240231[db BAA74894.1] - (AB020678) KIAA0871 protein [Homo sapiens]		struct	56182575, 29331822, 29331824, 29331827, 66712502, 264591, 33657402, 60433356, 265019, 21906768, 21906769, 35695917, 265020, 265021, 264636, 56182323
1733	87617178 (3465, 3466)	Novel Protein sim. GBank gij1575756 (U70674) - m-Numb [Mus musculus]	Contains protein domain (PF00640) - synthase Phosphotyrosine interaction domain (PTB/PID).		264907, 264910, 33657402, 265010, 264681, 264693, 264684, 264686, 264769, 264691, 264692, 264693, 264628, 264636, 264556
1734	87785261 (3467, 3468)				264693



1735	88318638 (3469, 3470)	Novel Protein sim. GBank gi 4836807 gb AAD30566.1 AF14679 - (AF146793) PFT27 [Mus musculus]	Contains protein domain (PF01169) - Uncharacterized protein family UPF0016	264488, 18108394, 18108398, 22278998, 60432049, 264259, 29331822, 264908, 265006, 265007, 265008, 265009, 264591, 33657402, 265010, 265011, 87168559, 265017, 18108351, 264682, 18108354, 264769, 264689, 21906765, 21906768, 21906769, 29148629, 29148784, 265021, 265022, 52644150, 18108364, 18108365, 33657109, 18108370, 18108374, 18108380, 18108385, 87168518, 264563, 18108390, 264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1736	95362884 (3471, 3472)	Novel Protein sim. GBank gi 4885647 ref NP_005472.1 pTRAP - thyroid hormone receptor-associated protein complex component	UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1737	88165549 (3473, 3474)	Novel Protein sim. GBank gi 2143607 pir S68695 - B/K protein - rat	Contains protein domain (PF00169) - C2 domain	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1738	85788811 (3475, 3476)	Novel Protein sim. GBank gi 2225941 emb CAA69714  - (Y08460) Mdes protein [Mus musculus]	UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1739	87328576 (3477, 3478)		UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1740	83592939 (3479, 3480)	Novel Protein sim. GBank gi 4809 emb CAA44309  - (X62452) YCR601 [Saccharomyces cerevisiae]	Contains protein domain (PF00441) - Acy-CoA dehydrogenase	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1741	95010100 (3481, 3482)	Novel Protein sim. GBank gi 4883898 gb AAD31695.1 AF13042 - (AF130420) serine protease-like protein isoform [Homo sapiens]	Contains protein domain (PF01363) - FYVE zinc finger	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1742	85788814 (3483, 3484)	Novel Protein sim. GBank gi 4505193 ref NP_003667.1 pMLD - membrane fatty acid (lipid) desaturase	UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1743	86966475 (3485, 3486)	Novel Protein sim. GBank	UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1744	91224003 (3487, 3488)	gi 728832 sp P39189 ALU2_HUMAN - !!!!! ALU SUBFAMILY SB WARNING ENTRY !!!!!	UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1745	20290075 (3489, 3490)	Novel Protein sim. GBank	UNCLASSIFIED	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1746	94326110 (3491, 3492)	Novel Protein sim. GBank gi 731756 sp P38873 YHY6_YEAST - HYPOTHETICAL 175.8 KD PROTEIN IN GND1-1K1 INTERGENIC REGION	Contains protein domain (PF00400) - WD domain, G-beta repeat	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564
1747	94324333 (3493, 3494)	Novel Protein sim. GBank gi 1658503 (U75467) - Alu [Drosophila melanogaster]	transcript factor	264490, 264259, 66714117, 66712502, 56182435, 265006, 265008, 264910, 60433358, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264564

1748	88003580 (3495, 3496)	Novel Protein sim. GBank gil4504511[ref]NP_001530.1[phsJ2 - heat shock protein, DNAJ-like 2]	Contains protein domain (PF00684) - DnaJ central domain (4 repeats)	eph	264489, 56182575, 29331824, 56182435, 264112, 265007, 265019, 264764, 21906768, 265020, 264691, 55811576, 264635, 264555, 264556, 264557, 264559 264106
1749	83363091 (3497, 3498)	Novel Protein sim. GBank gil5650780[gb]AAD45948.1[AF15196 - (AF151968) RGS protein RGS-17 (Gallus gallus)]	Contains protein domain (PF00615) - Regulator of G protein signaling domain	oncogene	
1750	94321664 (3499, 3500)	Novel Protein sim. GBank gil4996894[gb]AAC28444.2] - (AF065164) hyperpolarization-activated, cyclic nucleotide- gated channel 2 [Homo sapiens]			33657402, 264288, 52644150, 263974, 83373044
1751	83373058 (3501, 3502)	Novel Protein sim. GBank gil276016[dbj]BA24184] - (AB010054) outer arm dynein light chain 2 [Anthriscus crassispina]	Contains protein domain (PF00560) - Leucine Rich Repeat	ATPase associated	265010, 264369
1752	86456530 (3503, 3504)	Novel Protein sim. GBank gil3915482[sp]P74346[YG29 - SYN3 - HYPOTHETICAL 35.0 KD PROTEIN SLR1629]	Contains protein domain (PF00849) - RNA pseudouridylylase synthase	deaminase	264510, 264593, 264682, 21906765, 18108370
1753	94235159 (3505, 3506)	Novel Protein sim. GBank gil2852636 (AF007155) - unknown [Homo sapiens]	Contains protein domain (PF01553) - Acyltransferase	phosphatase	56994075, 22278996, 264908, 60170831, 264682, 264784, 264369, 264288, 264685, 264687, 21906766, 264692, 264693 65274620, 65274791, 35695855, 264637, 264564
1754	88095323 (3507, 3508)	Novel Protein sim. GBank gil731421[sp]P39881[YEHA - YEAST - HYPOTHETICAL 53.3 KD PROTEIN IN HXT8-CAN1 INTERGENIC REGION]		transport	264488, 35696286, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 264591, 33657402, 264594, 264757, 264758, 264600, 264604, 264762, 264763, 264685, 264766, 264691, 264628, 35696423, 264632, 264634, 264637, 264638, 263981, 264558, 264639, 264563, 264564, 264565, 264566, 264567 264686
1755	79470282 (3509, 3510)	Novel Protein sim. GBank gil1176422 (U43194) - rhophilin [Mus musculus]		UNCLASSIFIED	
1756	92962614 (3511, 3512)	Novel Protein sim. GBank gil4432860[gb]AAD20708] - (AC006300) putative glucose-induced repressor protein [Arabidopsis thaliana]			52648842, 22278994, 22278995, 56994075, 22278996, 22278997, 264259, 29331822, 60432289, 29331827, 33656970, 265006, 265009, 60432229, 60433356, 60433438, 33109954, 21906754, 265017, 265018, 265019, 264448, 264369, 264288, 21906765, 21908768, 21908767, 21908768, 21908769, 265020, 265021, 265022, 264692, 27486262, 27486264, 18108376, 20281152, 264558, 18108388, 87168516, 60432113, 22279000, 22279002, 264482 22278997, 264259, 60432289, 29331827, 264908, 52644045, 265008, 264593, 265019, 264766, 21906768, 65274620, 18108385, 60432113, 264566, 264487
1757	95357380 (3513, 3514)	Novel Protein sim. GBank gil5441615[emb]CAB46856.1] - (AJ388557) zinc finger protein [Canis familiaris]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_rna_bind	

1758	87612971 (3515, 3516)	Novel Protein sim. GBank gij3881040[emb]CAA164031 - (AL021497) predicted using Genefinder [Caenorhabditis elegans]		UNCLASSIFIED	22278994, 22278999, 29331822, 29331824, 29331825, 33656970, 264508, 265006, 265007, 265009, 264591, 33657402, 33109954, 87168474, 264600, 265017, 265018, 21906769, 265020, 265021, 33657023, 33657109, 264629, 18108374, 35695855, 264632, 52644332, 22279002, 264563
1759	36994372 (3517, 3518)			UNCLASSIFIED	264759
1760	87329716 (3519, 3520)	Novel Protein sim. GBank gij5262748[emb]CAB45688.11 - (AJ133120) Proline rich synapse associated protein 2 [Rattus norvegicus]		UNCLASSIFIED	56182575, 60432049, 35696052, 264905, 264906, 264907, 264908, 264909, 265006, 265009, 264910, 60432228, 264592, 264595, 55812038, 264758, 264762, 18108351, 264764, 264766, 264768, 264769, 21906765, 55811957, 35695917, 264690, 264692, 264628, 264629, 55811576, 35696423, 264632, 264634, 264636, 264557, 264639, 60432113, 264565, 264486
1761	87409588 (3521, 3522)	Novel Protein sim. GBank gij127749[sp]P10569[MYSC_ACACA - MYOSIN IC HEAVY CHAIN	Contains protein domain (PF00618) - Guanine nucleotide exchange factor for Ras-like GTPases; N-terminal motif	UNCLASSIFIED	29331822, 264910, 264685, 264686
1762	95319887 (3523, 3524)	Novel Protein sim. GBank gij3169158 (AC004770) - BC289730_2 [Homo sapiens]	Contains protein domain (PF00173) - Heme-binding domain in cytochrome b5 and oxidoreductases	desaturase	22278999, 264259, 264905, 264907, 60170831, 265010, 265011, 265017, 264448, 21906765, 21906766, 21906767, 21906768, 265021, 264690, 33657109, 18108374, 264558, 60170394
1763	91224013 (3525, 3526)	Novel Protein sim. GBank gij4809026[gbl]AAD30062.11 - (AF132856) suppressor of G2 allele of stk1 homolog [Homo sapiens]			56181686, 28331825, 35696052, 264905, 264906, 264909, 264763, 264682, 264769, 35695917, 265022, 33657023, 18108374, 35696423, 264634
1764	87757687 (3527, 3528)	Novel Protein sim. GBank gij1360669[pilr]CGHU1V - collagen alpha 1(V) chain precursor - human	Contains protein domain (PF01410) - Fibrillar collagen C-terminal domain	collagen	22278998, 264259, 35696052, 29331828, 56182435, 265008, 265017, 265018, 264448, 264288, 21906766, 21906767, 29148627, 35695917, 264691, 33657023, 60432113, 22279002
1765	91230091 (3529, 3530)	Novel Protein sim. GBank gij486806[pilr]S35503 - finger protein neutralized - fruit fly (Drosophila melanogaster)		UNCLASSIFIED	264488, 18108394, 22278995, 22278997, 22278999, 264259, 264104, 264508, 264905, 264807, 264511, 265007, 265008, 264910, 265009, 21906754, 265010, 265017, 264603, 265018, 265019, 18108351, 264682, 264448, 264369, 264288, 264766, 18108359, 21906766, 21906767, 29148627, 29148629, 35695917, 265020, 265021, 264692, 264628, 264629, 18108374, 263976, 264636, 83373044, 22279000, 22279002

1766	95081201 (3531, 3532)	Novel Protein sim. GBank gi 2499087 sp Q09332 UGG_DROME - UDP- GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT)		glycoprotein	52645156, 87168559, 60170815, 33657023, 264693, 33657109, 27486261, 264555, 83373044
1767	87755998 (3533, 3534)	Novel Protein sim. GBank gi 4176443 emb CAA18263.1  - (AL022238) dJ1042K10.4 (novel protein) [Homo sapiens]		UNCLASSIFIED	265017, 265019, 264686, 264768, 265020, 264692
1768	80253216 (3535, 3536)			UNCLASSIFIED	29331824, 29331825, 264591, 56182323
1769	87388988 (3537, 3538)			UNCLASSIFIED	264563
1770	85413144 (3539, 3540)	Novel Protein sim. GBank gi 115204 sp P00736 C1R_HUMAN - COMPLEMENT C1R COMPONENT PRECURSOR	Contains protein domain (PF00089) - Trypsin	complement	264488, 264768, 264769, 56182575, 55811957, 264690, 264691, 35696052, 264905, 264509, 264906, 264907, 264628, 264908, 264909, 264910, 264634, 264635, 264636, 264556, 264757, 264758, 55812038, 65274444, 264760, 264563, 264762, 264764, 264684, 264766
1771	94233542 (3541, 3542)	Novel Protein sim. GBank gi 3914191 sp P56558 OGT1_RAT - UDP-N- ACETYLGLUCOSAMINE-PEPTIDE N- ACETYLGLUCOSAMINYLTRANSFERASE 110 KD SUBUNIT (O-GLCNAC TRANSFERASE P110 SUBUNIT)	Contains protein domain (PF00515) - TPR Domain	transferase	264758, 264600, 264369, 55811957, 265020, 83373044, 22279000
1772	87643510 (3543, 3544)	Novel Protein sim. GBank gi 4859442 gb AAD34351.1 AF12136 - (AF121360) DNZDHH/C/NEW1 zinc finger protein 11 [Drosophila melanogaster]	Contains protein domain (PF01529) - DRHC zinc finger domain		22278998, 29331828, 33109954, 265018, 265019, 264764, 21906765, 265020, 265021, 264558
1773	94116824 (3545, 3546)	Novel Protein sim. GBank gi 3978464 (AF085693) - G protein-coupled receptor kinase-associated ADP ribosylation factor GTPase-activating protein [Rattus norvegicus]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	potassium_channel	65274572, 56182575, 22278996, 35696286, 22278999, 264259, 29331824, 60424269, 29331825, 60432289, 35696052, 264106, 264509, 264906, 264907, 29331830, 264908, 52644045, 264511, 265006, 265007, 265008, 60170831, 60433438, 264758, 55811386, 87168559, 265017, 264604, 265019, 55811150, 264288, 56181562, 264689, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265022, 52644150, 264691, 33657023, 264692, 264693, 60431528, 35696423, 35695855, 264636, 56182323, 18108387, 56526486, 22279000, 22279002, 264563, 264564, 264565, 264566, 264567
1774	94232573 (3547, 3548)	Novel Protein sim. GBank gi 2495699 sp Q15034 Y032_HUMAN - HYPOTHETICAL PROTEIN KIAA0032	Contains protein domain (PF00415) - Regulator of chromosome condensation (RCC1)	UNCLASSIFIED	65274572, 56182575, 35696052, 55812038, 33109954, 21906754, 265017, 21906767, 21906768, 21906769, 265020, 264691, 264636, 56182323, 22279002

1775	95359330 (3549, 3550)	Novel Protein sim. GBank gi 1469199 dbj BAA09487  - (D50928) The KIAA0138 gene product is novel. [Homo sapiens]		UNCLASSIFIED	65274572, 56182575, 56994075, 35696286, 22278997, 29331822, 29331826, 60432289, 29331828, 35696052, 29331830, 68712502, 264828, 56182435, 264511, 265007, 265009, 60170831, 60432229, 60433438, 55812038, 21906754, 85658542, 87168559, 264601, 265017, 265018, 265019, 264762, 264448, 264288, 264689, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 265022, 264691, 33657023, 264692, 33657109, 27486261, 33657349, 18108370, 18108377, 35696423, 55811576, 35695855, 264632, 264634, 264636, 264639, 56182323, 83373044, 56526486, 87168518, 60432113, 22279000, 22279002, 264482, 264486, 264910
1776	94133756 (3551, 3552)	Novel Protein sim. GBank gi 4589676 dbj BAA76857.1  - (AB023230) KIAA1013 protein [Homo sapiens]			
1777	87447171 (3553, 3554)	Novel Protein sim. GBank gi 3219939 sp P87115 YDK9_SCHPO - HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME 1		nuc1_rept	56994075, 29331826, 265008, 87168474, 265017, 265018, 264761, 55811150, 264764, 56181562, 264689, 21906765, 21906768, 21906769, 35695917, 264690, 33657023, 35695763, 60431528, 35696423, 55811576, 35695855, 22279000, 22279002, 264584
1778	94851624 (3555, 3556)	Novel Protein sim. GBank gi 3675648 emb CAA91454.1  - (Z88561) Similarity to Human rab13 protein (PIR Acc. No. A49647). Contains the ATP/GTP-binding site motif (PROSITE PS00017); cDNA EST EMBL:M89412 comes from this gene; cDNA EST yk212g9.3 comes from this gene; cDNA EST yk212g9...		UNCLASSIFIED	29331826, 29331827, 35696052, 264512, 265007, 265009, 265017, 265019, 264762, 18108351, 264769, 21906765, 21906768, 21906767, 21906768, 21906769, 265020, 265021, 264691, 264693, 18108370, 264556, 83373044
1779	94133758 (3557, 3558)	Novel Protein sim. GBank gi 4589676 dbj BAA76857.1  - (AB023230) KIAA1013 protein [Homo sapiens]		UNCLASSIFIED	29148627, 35696286, 29147620, 265006, 265007, 265008, 18108385, 65274727, 264482, 264369, 264766
1780	87023497 (3559, 3560)		Contains protein domain (PF00807) - Apidaecin	UNCLASSIFIED	264107, 33657109, 56526486
1781	84047477 (3561, 3562)			UNCLASSIFIED	264508, 264906, 264639
1782	88094607 (3563, 3564)	Novel Protein sim. GBank gi 729225 sp P41237 CTXN_RAT - CORTEXIN			264259, 28331822, 264508, 264905, 264906, 264907, 264908, 265007, 265009, 264910, 264591, 264758, 264764, 264288, 264768, 264769, 264635, 264636, 264637, 264639, 264563
1783	85717805 (3565, 3566)	Novel Protein sim. GBank gi 2257543 dbj BAA21436  - (AB004538) protein arginine N-methyltransferase [Schizosaccharomyces pombe]		interferon	264768

1784	95197093 (3567, 3568)	Novel Protein sim. GBank gj1755048 (U55042) - myosin X [Bos taurus]	Contains protein domain (PF00169) - struct PH domain	35696286, 264259, 35696052, 264508, 264905, 264906, 264907, 66712502, 264908, 264909, 265007, 265008, 265009, 264910, 264591, 264594, 264757, 264758, 264759, 265010, 265011, 264601, 264602, 264604, 264605, 18108351, 264762, 264763, 264764, 264369, 264766, 264687, 264768, 264688, 21906768, 35695917, 264690, 264691, 264692, 264693, 264628, 18108374, 35696423, 264631, 264632, 264635, 264637, 264638, 264639, 18108385, 22279000, 22279002, 264565, 264566, 264486
1785	95357475 (3569, 3570)	Novel Protein sim. GBank gj14589552(dbj BAA76798.1) - (AB023171) KIAA0954 protein [Homo sapiens]	UNCLASSIFIED	65274572, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 264905, 264906, 264908, 66712502, 56182435, 264511, 265007, 60433356, 55811150, 264683, 264369, 264687, 52644229, 21908767, 52644150, 33657023, 65274620, 33657182, 65274791, 35695855, 264555, 35274727, 22279002
1786	85286465 (3571, 3572)	Novel Protein sim. GBank gj117788 sp P26770 CYA4_RAT ADENYLATE CYCLASE, TYPE IV (ATP PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE)	UNCLASSIFIED	264908, 35696423, 264636
1787	87434784 (3573, 3574)	Novel Protein sim. GBank gj13877175 emb CAA90338.1  - (Z50028) cDNA EST yk321h8.5 comes from this gene; cDNA EST EMBL:D88896 comes from this gene; cDNA EST yk395f9.5 comes from this gene [Caenorhabditis elegans]	UNCLASSIFIED	264488, 264905, 264908, 264909, 264595, 264764, 264766, 264692, 60431528, 264629, 264636, 264584, 264566
1788	91228779 (3575, 3576)		UNCLASSIFIED	264488, 83373044
1789	88094529 (3577, 3578)	Novel Protein sim. GBank gj12088669 (AF003130) - F55A12.9 gene product [Caenorhabditis elegans]	UNCLASSIFIED	264488, 29331828, 264909, 18108351, 264288, 265021, 264555, 264636
1790	82489734 (3579, 3580)			35696052, 264905, 264906, 264907, 264908, 264909, 265008, 264910, 264758, 265011, 265019, 264764, 264766, 264769, 264628, 264635

1781	95197259 (3581, 3582)	Novel Protein sim. GBank gij2114321[dbj BAA200371] - (D88733) membrane glycoprotein [Equine herpesvirus 1]	Contains protein domain (PF00047) - Immunoglobulin domain	glycoprotein	264488, 264686, 264687, 264768, 18108394, 264769, 18108397, 264259, 264691, 264692, 33657023, 264693, 264509, 264905, 264906, 264628, 264907, 264629, 264908, 264909, 264510, 265006, 264511, 265008, 264630, 265009, 264631, 264910, 264632, 264634, 264635, 264555, 264636, 264592, 264637, 264593, 264638, 18108381, 264639, 264758, 265010, 265011, 264602, 22279000, 264604, 264760, 264584, 264681, 264762, 264565, 264763, 264683, 264566, 264764, 264288, 264584, 264567, 18108354, 18108391, 264685, 264766
1782	87792690 (3583, 3584)	Novel Protein sim. GBank gij4337106[gib AAD18082] - (AF129758) BAT4 [Homo sapiens]	Contains protein domain (PF01585) - G-patch domain	UNCLASSIFIED	22278997, 264259, 264508, 265007, 33657402, 87168559, 264369, 33657023, 35695855, 20281071, 264559, 18108387, 87168518
1783	85337877 (3585, 3586)	Novel Protein sim. GBank gij5579331[gib AAD45504.1 AF145732] endoplasmic reticulum alpha-mannosidase 1 [Homo sapiens]	Contains protein domain (PF01532) - Glycosyl hydrolase family 47	ATPase associated	65274572, 22278995, 22278996, 22278997, 22278998, 264093, 264259, 28331824, 66714117, 60432289, 28331827, 29331828, 284103, 264105, 28331830, 265007, 264910, 265009, 60170831, 60433356, 21906754, 265010, 265017, 265019, 264681, 264682, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 60170615, 52844150, 33657023, 33657109, 18108370, 18108374, 65274791, 20281071, 60432113, 22279000, 264482, 264564
1784	87759808 (3587, 3588)	Novel Protein sim. GBank gij4814604[embj CAB43677.1] - (AL050369) hypothetical protein [Homo sapiens]	Contains protein domain (PF01798) - Putative snoRNA binding domain	UNCLASSIFIED	18108394, 22278995, 22278999, 264259, 29331822, 29331824, 29331825, 29146498, 29146499, 264508, 264905, 52844045, 264112, 265006, 265008, 264910, 60433356, 264757, 55812038, 87168474, 265011, 265017, 18108351, 264763, 264448, 264683, 264369, 21906765, 21906766, 21906767, 21906769, 29148794, 35695917, 60170615, 33657023, 264629, 18108374, 18108376, 35696423, 35695855, 264556, 264557, 264638, 264558, 18108385, 264564
1785	79747856 (3589, 3590)			UNCLASSIFIED	264632, 264635, 264636, 264595, 264596, 264907, 264566, 264909
1786	86589486 (3591, 3592)	Novel Protein sim. GBank gij585084[sp Q07803 EFGM_RAT - ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G)]		glycoprotein	264488, 264907, 264909, 264594, 264595, 264766, 264687, 21906765, 21906767, 264628, 264630, 264559

1797	91223219 (3593, 3594)	Novel Protein sim. GBank gi 1842111 (U87586) - decoy [Arabidopsis thaliana]		ribosomalprot	22278996, 22278997, 22278998, 22278999, 29331822, 264910, 60170831, 21906754, 52644229, 21906765, 21906768, 21906769, 35695917, 265022, 52644150, 264691, 33657023, 263967, 33657109, 22279000
1798	91221276 (3595, 3596)	Novel Protein sim. GBank gi 2832806 dbj BAA24608.1  - (D89340) dipeptidyl peptidase III [Rattus norvegicus]		peptidase	22278994, 56994075, 22278997, 22278998, 22278999, 264259, 29331826, 60432289, 29331828, 33656970, 265008, 60432229, 264757, 60433438, 21906754, 33657084, 87168559, 265017, 18108351, 264682, 264448, 264288, 21906765, 21906768, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 33657023, 33657182, 27486261, 27486265, 33657349, 263973, 18108374, 55811576, 35695855, 18108385, 87168518, 22279000, 264486
1799	86321713 (3597, 3598)	Novel Protein sim. GBank gi 5689541 dbj BAA83054.1  - (AB029025) KIAA1102 protein [Homo sapiens]		eph	264908, 21906754, 21906767, 21906769, 265020, 33657023, 264692, 264693, 264404, 22279000
1800	87080116 (3599, 3600)	Novel Protein sim. GBank gi 4880679 gb AAD27729.1 AF13295 - (AF132954) CGI-20 protein [Homo sapiens]		UNCLASSIFIED	264691, 264556, 264566
1801	95060723 (3601, 3602)	Novel Protein sim. GBank gi 4880679 gb AAD27729.1 AF13295 - (AF132954) CGI-20 protein [Homo sapiens]			52644045, 265007, 264632
1802	87771012 (3603, 3604)	Novel Protein sim. GBank gi 134920 sp P21997 SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)			35696286, 66714117, 264508, 264509, 56182435, 264512, 18108351, 264688, 55811957, 264692, 55811576, 35695855, 264486
1803	95060725 (3605, 3606)	Novel Protein sim. GBank gi 4880679 gb AAD27729.1 AF13295 - (AF132954) CGI-20 protein [Homo sapiens]			264686, 264488, 264687, 264489, 264768, 264769, 264689, 21906769, 35696286, 35695917, 264259, 264691, 264692, 264693, 20281098, 18108364, 35696052, 264508, 264509, 264905, 264906, 18108370, 264628, 264907, 66712502, 264908, 264909, 18108374, 18108376, 35696423, 35695855, 264510, 264511, 265006, 265007, 264512, 265008, 264910, 264631, 264632, 264634, 264635, 264591, 264636, 264637, 264592, 264638, 264593, 264639, 264594, 83373044, 264758, 264598, 18108385, 18108387, 265011, 264760, 264563, 18108351, 264762, 264564, 264448, 264565, 264763, 264683, 264764, 264586, 264288, 264486, 264567, 264765, 264766



1804	87770203 (3607, 3608)	Novel Protein sim. GBank gjl38789914[emb]CAA98538.1] - (Z74043) predicted using GeneFinder. cDNA EST EMBL:C13850 comes from this gene: cDNA EST EMBL:C11575 comes from this gene: cDNA EST yk34314.5 comes from this gene [Caenorhabditis elegans]			52645355, 22278997, 22278999, 264905, 264908, 264909, 264910, 21906754, 264766, 21906765, 21906768, 35695917, 265020, 265022, 264691, 264637, 264639, 22279000, 264564, 264566
1805	95330375 (3608, 3610)	Novel Protein sim. GBank gjl5453644[ref]NP_006461.1pEBBP - estrogen-responsive B box protein			29331824, 29331825, 29331826, 29331827, 29331828, 87168559, 264288, 264687, 52644229, 35695423, 264638, 60432113
1806	94133762 (3611, 3612)	Novel Protein sim. GBank gjl4589676[dbj]BAA76857.1] - (AB023230) KIAA1013 protein [Homo sapiens]	sinud		264094, 264105, 264908, 35696423, 265006, 265007, 265008, 264555, 264592, 265011, 265018, 264369
1807	86943032 (3613, 3614)				29331824, 264908, 264910, 33657023, 263978
1808	87642711 (3615, 3616)	Novel Protein sim. GBank gjl4884070[emb]CAB43235.1] - (AL050008) hypothetical protein [Homo sapiens]	UNCLASSIFIED		264488, 35696286, 66714117, 35696052, 66712502, 264592, 60433438, 52644296, 265010, 264683, 264369, 264689, 55811957, 35695917, 33657109, 35695763, 55810784, 18108379, 35696423, 35695855, 56182323, 264583, 264584, 264487
1809	95321468 (3617, 3618)	Novel Protein sim. GBank gjl1916027 (U87965) - putative G protein [Mus musculus]	UNCLASSIFIED		264594, 55811150, 264686, 29148629, 29148784, 264690, 264629, 18108374, 264556, 264557, 264558
1810	88086316 (3618, 3620)	Novel Protein sim. GBank gjl1352944[sp]P47179YJ9P_YEAST - HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR	UNCLASSIFIED		264488, 35696052, 264905, 264906, 264907, 264908, 264909, 264511, 265009, 264910, 264592, 264593, 264594, 33657402, 264757, 264595, 264758, 264596, 264759, 264600, 264601, 264762, 264683, 264764, 264288, 264684, 264766, 264767, 264686, 264768, 264687, 264769, 264689, 265021, 264690, 264691, 264693, 264628, 264629, 18108374, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 264563, 264566, 264486, 264567
1811	88086272 (3621, 3622)	Novel Protein sim. GBank gjl2134984[pir]j137275 - death-associated protein kinase (EC 2.7.1.-) - human	kinase Ank repeat	Contains protein domain (PF00023) - Ank repeat	264488, 264259, 264508, 264509, 264905, 264906, 264907, 56182435, 264511, 264512, 264910, 264758, 265011, 264600, 264604, 18108354, 264766, 264686, 264769, 264534, 60170615, 33657023, 264629, 264631, 264639, 264583, 264482, 264483
1812	79245772 (3623, 3624)				29331822, 29331824, 265019, 18108351, 21906769

1813	88090972 (3625, 3626)	Novel Protein sim. GBank gi15051636 gi A038326.1 AF07372.2 - (AF073727) EH domain-binding mitotic phosphoprotein [Homo sapiens]	Contains protein domain (PF01417) - ENTH domain	glucoamylase	56182575, 264259, 29331824, 66714117, 29331828, 35696052, 264509, 264905, 264908, 264907, 264908, 66712502, 264909, 265007, 264910, 264991, 264593, 55812038, 265011, 265018, 264760, 264882, 264784, 264683, 264369, 264768, 264768, 264769, 21906766, 21906788, 264691, 264693, 18108374, 35695855, 264634, 264635, 264637, 264639, 264559, 22279000, 22279002, 264566
1814	88178047 (3627, 3628)	Novel Protein sim. GBank gi13643608 (AC005395) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 35696266, 22278998, 264092, 264094, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 264105, 264107, 52644045, 56182435, 265009, 60432229, 60433356, 87168474, 87168559, 264369, 264288, 21906765, 35695917, 265021, 265022, 33657023, 33657109, 18108374, 35696423, 264638, 56526486, 264482
1815	85296473 (3629, 3630)	Novel Protein sim. GBank gi117788 sp P26770 CYA4_RAT ADENYLATE CYCLASE, TYPE IV (ATP PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE)		struc1	22278999, 264508, 264509, 264907, 264908, 264910, 265011, 264760, 264766, 264634, 264636
1816	83738845 (3631, 3632)	Novel Protein sim. GBank gi1176623 sp P41846 YO96_CAEEL - HYPOTHETICAL 93.9 KD PROTEIN T20812.6 IN CHROMOSOME III		UNCLASSIFIED	18108394, 18108397, 264509, 264907, 264908, 264909, 265008, 264591, 265011, 265017, 264687, 264688, 265022, 264691, 18108362, 18108368, 18108370, 18108374, 18108379, 264635, 264557, 264564, 264567
1817	88095268 (3633, 3634)	Novel Protein sim. GBank gi13766377 emb CAA21429  - (AL031907) hypothetical protein [Schizosaccharomyces pombe]	Contains protein domain (PF00400) - WD domain, G-beta repeat	histone	264488, 22278997, 22278999, 60432049, 29331822, 29331824, 60432289, 52644045, 60170831, 265017, 265018, 265019, 18108351, 264682, 52644229, 21906765, 21906767, 21906768, 52644150, 33657023, 33657109, 27486262, 18108370, 18108374, 60170394, 56182323, 22279002
1818	85806775 (3635, 3636)	Novel Protein sim. GBank gi1387912 emb CAA94370  - (Z70310) predicted using GeneFinder: Similarity to Mouse ankyrin (PIR Acc. No. S37771); cDNA EST EMBL: T01923 comes from this gene; cDNA EST EMBL: D32335 comes from this gene; cDNA EST EMBL: D32723 comes from this gene; cDNA EST E...	Contains protein domain (PF00023) - Ank repeat	transcriptfactor	35696286, 60433356, 264758, 264369, 264686, 21906769, 264693, 264632
1819	87759572 (3637, 3638)	Novel Protein sim. GBank gi15031865 ref NP_005771.1 pLHFP - lipoma HMGIC fusion partner		UNCLASSIFIED	22278995, 22278997, 22278999, 264259, 29331826, 264908, 265007, 265008, 265009, 60432229, 265017, 265018, 265019, 264448, 264288, 21906768, 21906769, 265020, 18108381, 18108384, 22279000, 22279002, 264567

1820	87769455 (3639, 3640)				264905, 264907, 264594
1821	80431510 (3641, 3642)				264907, 264768, 263978
1822	91221523 (3643, 3644)	Novel Protein sim. GBank gll4884130[emb]CAB43272.1] - (AL050101) hypothetical protein [Homo sapiens]			22278995, 56994075, 22278996, 22278997, 22278998, 264259, 29331824, 29331825, 29331826, 35696052, 29331828, 264908, 29331830, 60170831, 264591, 264593, 60433356, 264596, 265017, 265019, 18108351, 264763, 264683, 21906765, 21908767, 21908768, 21906769, 35695917, 265020, 265021, 33657023, 18108364, 18108370, 35695855, 22279000, 22279002
1823	85522330 (3645, 3646)			UNCLASSIFIED	264488, 264259, 264511, 264288, 264768, 264693, 35696423, 264634, 18108385, 264486
1824	86612025 (3647, 3648)	Novel Protein sim. GBank gll477072[pr]IA48018 - mucin 7 precursor, salivary - human	Contains protein domain (PF00047) - Immunoglobulin domain	UNCLASSIFIED	264907, 264908, 264909, 264511, 264631, 264634, 264635, 264637, 264638, 264639, 264758, 264568
1825	87430125 (3649, 3650)	Novel Protein sim. GBank gll3036803[emb]CAA18493] - (AL022373) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	60432049, 264910, 264487
1826	91723612 (3651, 3652)	Novel Protein sim. GBank gll4680685[gb]AAD27732.1[AF132955] CGI-23 protein [Homo sapiens]		ATPase-associated	52644507, 52645156, 52646842, 22278994, 22278996, 56994075, 264259, 60432049, 52645080, 35696052, 66712502, 52644045, 265008, 265009, 60432229, 60433356, 60433438, 52646317, 52644296, 265011, 87168559, 264448, 264288, 264369, 264688, 52644229, 264689, 21906765, 21906768, 265020, 60170615, 52644150, 33657023, 27486262, 27486264, 27486265, 35695763, 35696423, 35695855, 83373044, 87168518, 264404, 22279002
1827	81647212 (3653, 3654)				264758

1828	95074017 (3655, 3656)	Novel Protein sim. GBank gi 4503571 ref NP_001419.1 pENO1 - endolase 1, (alpha)	Contains protein domain (PF00113) - oncogene Endolase	264488, 52646842, 56182575, 22278996, 35696286, 22278997, 22278998, 264091, 264093, 60432049, 264259, 29331822, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 264105, 264508, 264907, 66712502, 52644045, 56182435, 265006, 264511, 264512, 265007, 265008, 265009, 60170831, 60432229, 264593, 60433356, 60433438, 264758, 33109954, 21908754, 87168474, 265010, 265011, 87168559, 265017, 265018, 264761, 264762, 264448, 264784, 264883, 264288, 264369, 18108355, 264768, 18108357, 18108358, 264688, 264769, 264689, 21906768, 21906769, 35695917, 265021, 60170615, 33657023, 33657349, 263972, 55811576, 35695855, 264635, 264555, 264556, 264638, 264557, 87168518, 22279000, 22279002, 264563, 264482, 264555, 264484, 264567 264508, 264634, 264509, 264482, 29331827, 264908, 265009, 264910
1829	80187720 (3657, 3658)			
1830	94312942 (3659, 3660)	Novel Protein sim. GBank gi 2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma- associated herpesvirus]	nuclease	52645156, 22278994, 22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 29331822, 29331825, 35696052, 52646317, 52644296, 87168559, 265019, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 35695763, 263974, 35696423, 35695855, 52644332 29331824, 35696052, 29331830, 264595, 264758, 265010, 265019, 265022, 264693, 85274791 264602
1831	84138063 (3661, 3662)		UNCLASSIFIED	
1832	84521663 (3663, 3664)	Novel Protein sim. GBank gi 1330345 (U58755) - coded for by C. elegans cDNA yk34b1.5; coded for by C. elegans cDNA yk13h10.5; coded for by C. elegans cDNA yk46e8.5; coded for by C. elegans cDNA yk46d5.5; coded for by C. elegans cDNA yk43c2.5; coded for by C. elegans cDNA yk46e8....		

1833	85314184 (3665, 3666)	Novel Protein sim. GBank gi15174413 ref NP_006026.1 pCDC4 - CDC42-binding protein kinase beta (DMPK-like)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	22278994, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 60432289, 29331827, 35696052, 29146499, 264508, 264509, 264905, 264907, 66712502, 264908, 52644045, 264909, 264512, 265008, 264591, 264593, 60433356, 21906754, 33657084, 265011, 265017, 264604, 265018, 265019, 264681, 18108351, 264683, 264288, 264685, 264766, 264687, 21906768, 21906769, 29148629, 21906767, 21906768, 21906769, 29148629, 285020, 265021, 264690, 264692, 33657023, 65274620, 33657182, 27486284, 33657349, 65274791, 264634, 264635, 264556, 264557, 264558, 264559, 18108385, 58528486, 87188518, 60432113, 22279000, 22279002, 264563
1834	80562780 (3667, 3668)				264259, 264907, 264689, 22279000, 22279002
1835	94135718 (3669, 3670)			UNCLASSIFIED	22278998, 29331822, 29331826, 87168474, 264603, 21906768, 263976, 35695855, 83373044
1836	87348450 (3671, 3672)	Novel Protein sim. GBank gi14759286 ref NP_004268.1 pUCP4 - uncoupling protein 4	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	29331825, 264908, 265019, 264764, 264686, 21906765, 264635
1837	94234297 (3673, 3674)	Novel Protein sim. GBank gi13334400 sp Q24574 UBPE_DROME - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	22278995, 29146499, 265006, 265008, 265009, 265010, 264683, 21906765, 29148627, 29148629, 265020, 265021, 265022, 65274620, 18108370, 18108374, 264556, 18108385
1838	84324369 (3675, 3676)	Novel Protein sim. GBank gi11362599 p J A56154 - Abi substrate ena (enabled) - fruit fly (Drosophila melanogaster)	Contains protein domain (PF00568) - WH1 domain		28331822, 265017, 264760, 265020, 83373044
1839	87456508 (3677, 3678)	Novel Protein sim. GBank gi12117310 emb CAB09116.1  - (Z95620) hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	60433438, 264601, 21906765, 21906766, 265021, 33657109, 264556
1840	87391708 (3679, 3680)	Novel Protein sim. GBank gi127560 sp P23249 MV10_MOUSE - PROTEIN MOV-10		UNCLASSIFIED	264693
1841	85818445 (3681, 3682)	Novel Protein sim. GBank gi14572464 gb AA023834.1 AF12365 - (AF123653) FEZ1 [Homo sapiens]			56182575, 29331824, 29331826, 60433356, 264764, 264288, 33657023, 263967, 18108370, 18108374, 264631, 264555, 264556, 264639

1842	9092645 (3683, 3684)	Novel Protein sim. GBank gj1326268 (U58728) - C54H2.1 gene product [Caenorhabditis elegans]		UNCLASSIFIED	65274572, 29331822, 29331824, 29331825, 66714117, 29331826, 29331827, 264907, 264909, 52644045, 56182435, 264510, 265006, 265007, 265009, 264910, 60433356, 264757, 60433438, 55812038, 265017, 18108354, 264686, 264769, 33657023, 264693, 18108364, 33657109, 18108368, 264628, 55810764, 56182323, 18108384, 264563, 264564
1843	95292692 (3685, 3686)			UNCLASSIFIED	264488, 56182435, 264769, 29331826, 29331828, 264511, 265006, 265007, 264910, 264631, 264509, 264680, 264636, 264564, 264691, 60432229, 60432049, 264259, 264629, 33657023, 264486, 264909, 264567, 264595, 264766
1844	87444764 (3687, 3688)	Novel Protein sim. GBank gj12496887[sp]Q09232[YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C09F5.2 IN CHROMOSOME III]		UNCLASSIFIED	264908, 265022, 33657023, 87168518, 22279002
1845	95096673 (3689, 3690)	Novel Protein sim. GBank gj1175494[sp]Q09819[YAC5_SCHPO - HYPOTHETICAL 45.0 KD PROTEIN C16C9.05 IN CHROMOSOME I]	Contains protein domain (PF00628) - PHD-finger	transcriptfactor	264259, 29331824, 264907, 264908, 66712502, 264510, 265007, 265008, 55812038, 265018, 21908765, 52644150, 33657109, 264555, 264556, 264557, 56182323, 18108382, 83373044, 18108385, 264584
1846	84287872 (3691, 3692)	Novel Protein sim. GBank gj1388108[emb]CAA21739] - (AL032657) similar to EGF-like domain; cDNA EST yk298a12.3 comes from this gene; cDNA EST EMBL:D35398 comes from this gene; cDNA EST yk331h6.5 comes from this gene; cDNA EST yk299a12.5 comes from this gene; cDNA EST yk467g8.5 c...	Contains protein domain (PF00008) - EGF-like domain	UNCLASSIFIED	264905, 264908
1847	87821487 (3693, 3694)	Novel Protein sim. GBank gj15059323[gb]AAD38967.1[AF15152] hairy and enhancer of split related-1 [Homo sapiens]	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	transcriptfactor	22278997, 264259, 29331824, 264909, 18108351, 263974, 22279002
1848	86789360 (3695, 3696)	Novel Protein sim. GBank gj15701854[emb]CAB52191.1] - (AJ245417) G5b protein [Homo sapiens]		tm7	29331825, 29331826, 29331827, 265017, 264683, 264288, 264766, 264768, 21908767, 21908768, 264692, 22279002
1849	84287874 (3697, 3698)	Novel Protein sim. GBank gj14503665[ref]NP_001989.1[pFBLN - fibulin 2 precursor	Contains protein domain (PF00008) - EGF-like domain	ATPase_associated	56182575, 265018
1850	86689650 (3699, 3700)	Novel Protein sim. GBank gj14589582[dbj]BAA76813.1] - (AB023186) KIAA0989 protein [Homo sapiens]	Contains protein domain (PF00169) - PH domain	struct	60432049, 264908
1851	95418789 (3701, 3702)	Novel Protein sim. GBank gj1220637[dbj]BAA01477] - (D10627) zinc finger protein [Mus musculus]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_bind	29331824, 35696052, 264910, 60433438, 264688, 35695917, 265020, 52644150, 65274620, 52644332

1852	95413170 (3703, 3704)	Novel Protein sim. GBank gi 5174629 ref NP_006090.1 pPIAS - protein inhibitor of activated STAT3		UNCLASSIFIED	56182575, 35696286, 22278996, 22278997, 22278998, 264490, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 35696052, 52844045, 265007, 264910, 60432229, 60433358, 60433438, 55812038, 65274444, 265018, 265019, 18108351, 284448, 264686, 264687, 21906765, 21906767, 21906769, 265021, 265022, 52644150, 264693, 33657109, 18108370, 18108374, 55811576, 35695855, 56182323, 60432113, 22278002, 264583 264687, 264768, 52644507, 264769, 21906765, 21906767, 21906768, 22278995, 56994075, 22278999, 52844150, 264259, 264692, 29331822, 29331824, 52645129, 29331827, 33656970, 33657349, 35685763, 264508, 264906, 264628, 264907, 264629, 264909, 35696423, 35695855, 264510, 265006, 264511, 264512, 264630, 265009, 264631, 284910, 284634, 264635, 264637, 264593, 264638, 264639, 33657402, 18108385, 52646317, 52644296, 87168518, 87168559, 264602, 265017, 22279000, 265018, 264760, 264762, 264682, 264448, 264764, 264684, 264587, 264288, 264369, 264766 264592
1853	91222267 (3705, 3706)	Novel Protein sim. GBank gi 854065 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	
1854	86038152 (3707, 3708)	Novel Protein sim. GBank gi 2072964 (U83569) - putative p150 [Homo sapiens]		nuclease	
1855	91221459 (3709, 3710)	Novel Protein sim. GBank gi 4539520 emb CAB39994.1  - (AL035424) dA22D12.1 (novel protein similar to Drosophila Kelch (Ring Canal protein, KEL) and a heterogeneous set of other types of proteins) [Homo sapiens]	Contains protein domain (PF01344) - Kelch motif	nucd_rept	18108392, 52646365, 65274572, 56182575, 22278994, 56994075, 22278996, 35696286, 22278999, 264259, 60432049, 264908, 264510, 265007, 265008, 265009, 264595, 21906754, 87168474, 265011, 87168559, 264681, 264288, 264768, 18108359, 21906764, 21906768, 29148627, 265020, 265021, 265022, 52644150, 33657023, 33657109, 18108372, 18108374, 18108376, 35696423, 264631, 264636, 18108381, 264482
1856	94231871 (3711, 3712)	Novel Protein sim. GBank gi 3954978 emb CAA06945  - (AJ008278) acetylglucosaminyltransferase-like protein [Mus musculus]		UNCLASSIFIED	56994075, 264259, 29331828, 264511, 264910, 264758, 264693, 264637, 18108381, 83373044
1857	94324455 (3713, 3714)	Novel Protein sim. GBank gi 4322670 gb AAD16120  - (AF094508) dentin phosphoryn [Homo sapiens]		ATPase_associated	22278999, 264259, 264906, 60170831, 264448, 264686, 265020, 265022, 33657109, 60170394, 83373044

1858	87628311 (3715, 3716)	Novel Protein sim. GBank gi 4981903 gb AAD36415.1 AE00178 - (AE001788) ribosomal protein S15 [Thermotoga maritima]	Contains protein domain (PF00312) - Ribosomal protein S15	264757
1859	84407464 (3717, 3718)	Novel Protein sim. GBank gi 4240317 dbj BAA74937.1  - (AB020721) KIAA0914 protein [Homo sapiens]		22278986, 29331824, 265007, 33109954, 265019, 264369, 21906768, 29148784, 27486261, 52644332, 22279002 265019
1860	17929308 (3719, 3720)	Novel Protein sim. GBank gi 4009522 (AF099731) - connexin 31.1 [Homo sapiens]		
1861	88086370 (3721, 3722)	Novel Protein sim. GBank gi 2143637 pir j 84505 - calcium- dependent actin-binding protein - rat	Contains protein domain (PF00265) - Citrate synthase	264887, 264259, 29331822, 29331824, 29331825, 265007, 285009, 284591, 33109954, 265010, 265019, 264369, 264288, 264686, 264691, 264693, 27486264, 18108370, 18108374, 263977, 55811576, 56182323, 264639, 22279000, 22279002, 264482
1862	87372923 (3723, 3724)	Novel Protein sim. GBank gi 125493 sp P07313 KMLC_RABIT - MYOSIN LIGHT CHAIN KINASE, SKELETAL MUSCLE (MLCK)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	35696286, 264259, 87168474, 264369, 21906766, 264558, 264563
1863	85775037 (3725, 3726)	Novel Protein sim. GBank gi 3820909 emb CAA09299  - (AJ010842) Dof protein [Drosophila melanogaster]	UNCLASSIFIED	264601, 264766, 29148627, 29148629, 264692, 264629, 264635
1864	85547832 (3727, 3728)	Novel Protein sim. GBank gi 4322263 gb AAD15985  - (AF077738) metalloproteinase CPX-1 [Mus musculus]	Contains protein domain (PF00754) - F5/8 type C domain	22278999, 264259, 264907, 265018, 18108370, 264634, 264635, 264555, 264556, 284638, 18108387
1865	87740827 (3729, 3730)	Novel Protein sim. GBank gi 2495727 sp Q93073 Y256_HUMAN - HYPOTHETICAL PROTEIN KIAA0255		22278999, 264490, 29331822, 66714117, 66712502, 265006, 265007, 265008, 265009, 264591, 60433438, 265010, 265019, 264760, 264448, 264768, 29148627, 29148629, 265020, 265022, 18108385, 60432113
1866	87266816 (3731, 3732)	Novel Protein sim. GBank gi 5262617 emb CAB45748.1  - (AL080157) hypothetical protein [Homo sapiens]	kinase	18108374, 264769, 18108377, 21906765, 21906766, 35696423, 56182575, 21906769, 29148629, 35696286, 35695917, 265021, 264510, 264511, 264512, 264534, 264535, 60170831, 52644150, 264555, 264691, 264259, 264556, 264692, 264557, 33657023, 60433358, 29331822, 264559, 264595, 29331824, 18108385, 21906754, 33657182, 29331827, 35696052, 33656970, 87168518, 265017, 60431602, 22279000, 264508, 264509, 18108351, 264907, 264682, 264567, 18108372, 264765, 264486
1867	84579159 (3733, 3734)	Novel Protein sim. GBank gi 3859930 (AF078096) - forkhead/winged helix-like transcription factor 7 [Homo sapiens]	UNCLASSIFIED	264094



1868	87357459 (3735, 3736)	Novel Protein sim. GBank gi 3881525 emb CAA93884  - (Z70038) cDNA EST EMBL:D32579 comes from this gene; cDNA EST EMBL:D35254 comes from this gene; cDNA EST yk224b3.5 comes from this gene; cDNA EST yk357110.5 comes from this gene [Caenorhabditis elegans]		nuclease	264489, 22278997, 22278999, 29331825, 29331826, 265008, 265009, 33657402, 87168474, 18108351, 21906765, 21906768, 21906769, 265020, 265021, 60170615, 27486264, 264628, 18108374, 264631, 18108385, 87168518, 22279000, 22279002, 264566, 264567
1869	86977292 (3737, 3738)	Novel Protein sim. GBank gi 4826772 ref NP_004961.1 p IGFA - insulin-like growth factor binding protein, acid labile subunit	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264508, 264509, 264906, 264908, 264909, 264910, 264591, 264600, 18108351, 264683, 264766, 264769, 35695855, 264634, 264558, 264639, 18108385, 264583, 264488
1870	95349488 (3739, 3740)	Novel Protein sim. GBank gi 1869859 emb CAB06722  - (Z86099) very large tegument protein [human herpesvirus 2]		UNCLASSIFIED	29331824, 60424269, 265007, 265008, 21906754, 265017, 265018, 265019, 264288, 264766, 264686, 264688, 21906768, 21906769, 35695917, 60170815, 264692, 18108388, 35695763, 35698423, 65274791, 264638, 264639, 58526486
1871	80234464 (3741, 3742)			UNCLASSIFIED	264509, 264905, 264595, 264768, 264635, 264636, 264563, 264486
1872	80235355 (3743, 3744)	Novel Protein sim. GBank gi 2460316 (AF022147) - uterus-ovary specific putative transmembrane protein [Rattus norvegicus]		protease	264510, 264594, 264565
1873	80213890 (3745, 3746)				264509, 264512, 265009, 265011, 18108351, 264687, 264681, 18108370, 18108374, 264635
1874	95351136 (3747, 3748)		Contains protein domain (PF00293) - Bacterial mutT protein	UNCLASSIFIED	284488, 35695917, 264259, 264905, 264907, 264908, 264909, 263978, 264511, 264635, 264636, 264637, 264638, 33657402, 264558, 18108385, 264600, 264604, 264764, 264567, 264766
1875	87330516 (3749, 3750)	Novel Protein sim. GBank gi 4589520 dbj BAA76782.1  - (AB023155) KIAA0938 protein [Homo sapiens]		UNCLASSIFIED	35696286, 264828, 264592, 264557, 264558
1876	87112950 (3751, 3752)	Novel Protein sim. GBank gi 263810 bbs 22920 - collagen alpha chain [Riftia pachyptila=lube worms, Peptide, 1027 aa]		UNCLASSIFIED	264259, 29331822, 60432289, 264908, 264909, 264604, 264764, 264288, 264769, 18108376, 264556, 264558, 264559, 18108385
1877	87315208 (3753, 3754)	Novel Protein sim. GBank gi 3983356 gb AAC83924.1  - (AF102545) riboflavin binding protein precursor [Scaphiopus couchii]		UNCLASSIFIED	264767, 264686, 264768, 263978, 264693, 264639, 265010, 264563, 264905, 264908, 264907

1878	95351056 (3755, 3756)	Novel Protein sim. GBank gij4510345jgb AAD21434.1] - (AC006921) unknown protein [Arabidopsis thaliana]	Contains protein domain (PF01428) - ubiquitin AN1-like Zinc finger	264568, 264488, 35696286, 56994075, 264259, 29331822, 29331824, 29331825, 35696052, 29331828, 29148498, 264508, 264509, 264905, 264908, 264907, 264908, 264909, 264510, 264511, 264512, 265009, 264910, 33657402, 264594, 264758, 55812038, 265011, 264602, 264760, 18108351, 264762, 264681, 264682, 264764, 264389, 264288, 264766, 264687, 264768, 264769, 21906766, 35695917, 265021, 60170615, 33657023, 264692, 264693, 33657109, 27486265, 264628, 18108370, 264629, 18108374, 35696423, 264634, 264635, 264555, 264638, 264639, 83373044, 18108385, 56526486, 87168518, 264563, 264564, 264566, 264488, 264567
1879	95310883 (3757, 3758)	Novel Protein sim. GBank gij4928643jgb AAD34082.1 AF15184 - (AF151845) CGI-87 protein [Homo sapiens]	UNCLASSIFIED	264905, 264907, 264908, 265007, 264565, 264566
1880	91012978 (3759, 3760)	Novel Protein sim. GBank gij1550785jemb CAA69283] - (Y08026) Immune associated protein 38 [Mus musculus]	UNCLASSIFIED	264766, 264691, 264692, 83373044
1881	80214949 (3761, 3762)	Novel Protein sim. GBank gij93144jpir B40505 - hypothetical protein - suid herpesvirus 1 (strain Indiana-Funkhuser or Becker)	UNCLASSIFIED	264509, 264905, 264908, 264909, 264910, 264762, 264687, 33657023, 264632
1882	86582450 (3763, 3764)	Novel Protein sim. GBank gij2384956 (AF022985) - No definition line found [Caenorhabditis elegans]		264908, 21906766, 18108370, 263974, 87168518
1883	94216817 (3765, 3766)	Novel Protein sim. GBank gij1351218jpiP47226 TES2_MOUSE - TESTIN 2 (TES2) [CONTAINS: TESTIN 1 (TES1)]	Contains protein domain (PF00412) - LIM domain containing proteins	264908, 264910, 87168559, 21906766, 264636

1884	95310885 (3767, 3768)	Novel Protein sim. GBank gi 4929643 gb AAD34082.1 AF15184 - (AF151845) CGI-87 protein [Homo sapiens]		UNCLASSIFIED	264488, 18108394, 56181686, 22278998, 60432049, 264259, 29331822, 29331824, 29331825, 66714117, 29331827, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 56182435, 264510, 264511, 265007, 264512, 265009, 264910, 264592, 264593, 33657402, 55812038, 264758, 5581386, 265010, 265011, 87168559, 264600, 265017, 264604, 265019, 264605, 264760, 5581150, 264761, 264682, 264763, 264683, 264764, 264288, 264369, 264766, 264686, 264768, 264769, 29148784, 35695917, 264690, 264691, 33657023, 264692, 264693, 33657109, 18108370, 264628, 60431528, 264629, 263973, 18108374, 55810764, 55811576, 35696423, 35695855, 264631, 264634, 60431850, 264637, 264638, 56182323, 264639, 18108382, 83373044, 18108385, 60432113, 22279002, 264563, 264564, 264565, 264566, 264486, 264567
1885	87644280 (3769, 3770)	Novel Protein sim. GBank gi 2507155 sp P37370 VRP1_YEAST - VERPROLIN		UNCLASSIFIED	56182575, 264259, 264905, 264909, 265008, 264596, 264766, 265020, 264628, 60431528, 264634, 58526486, 264080, 264563
1886	86674062 (3771, 3772)	Novel Protein sim. GBank gi 2854158 gb AAC02577.1  - (AF045641) No definition line found [Caenorhabditis elegans]			22278998, 22278999, 60432049, 264910, 265018, 264766, 21906768, 29148629, 264690, 264693, 264628, 264555, 264488
1887	94139139 (3773, 3774)	Novel Protein sim. GBank gi 5174421 ref NP_006023.1 pCPNE - copine VI (neuronal)	Contains protein domain (PF00168) - C2 domain	ATPase-associated	29331822, 29331824, 29331825, 29331826, 29331827, 264906, 265007, 264681, 264768, 29148627, 264693, 18108364, 35696423, 65274791, 35695855, 264632, 56182323, 264639, 264563
1888	87822804 (3775, 3776)	Novel Protein sim. GBank gi 3319831 emb CAB10841 - (Z88046) dJ14O9.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens]	Contains protein domain (PF01454) - MAGE family		263978
1889	91255783 (3777, 3778)	Novel Protein sim. GBank gi 1083308 pir JAS6559 - enhancer-trap-locus-1 protein - mouse (fragment)	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	helicase	264906, 55812038, 264758, 265010, 265018, 265019, 18108351, 264288, 21906765, 21906768, 21906769, 265021, 33657023, 33657109, 56182323, 83373044, 18108385, 22279000, 22279002
1890	87626705 (3779, 3780)	Novel Protein sim. GBank gi 4240185 db BAA74876.1  - (AB020660) KIAA0853 protein [Homo sapiens]		UNCLASSIFIED	18108398, 29147620, 264907, 265009, 264600, 265018, 18108351, 264288, 264689, 21906765, 21906768, 21906769, 264691, 264692, 264693, 264628, 18108370, 264636, 264558, 264404

1891	87013895 (3781, 3782)			UNCLASSIFIED	264686, 264768, 264687, 264692, 264693, 29331822, 29331824, 264508, 264805, 264906, 18108370, 264628, 264907, 264908, 264909, 18108378, 265007, 265008, 264910, 264632, 264591, 264639, 264596, 18108384, 265010, 265011, 264601, 264605, 264563, 264369
1892	87642825 (3783, 3784)	Novel Protein sim. GBank gij5688535[dbj BAA83051.1] - (AB029022) KIAA1099 protein [Homo sapiens]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	struct	22278995, 264509, 87168559, 18108351, 264448, 264682, 265020, 264693, 18108374, 22279000
1893	88533826 (3785, 3786)			laminin	264569, 65274572, 22278987, 22278999, 264259, 29331822, 29331824, 68714117, 29331826, 264906, 265006, 265008, 265009, 264592, 265018, 264681, 264448, 264683, 18108354, 264368, 264684, 264685, 264766, 264687, 264689, 21906768, 265020, 265022, 60170615, 52644150, 264690, 264691, 264692, 33657023, 264693, 33657109, 264628, 18108374, 35695855, 264630, 264632, 264634, 264557, 264558, 60170394, 18108381, 18108385, 22279000
1894	86989120 (3787, 3788)				264508, 264905, 264908, 264907, 264594, 264684, 264690, 264692, 264630, 264635, 264636, 264639, 264563
1895	87831891 (3789, 3790)	Novel Protein sim. GBank gij5262574[emb CAB45729.1] - (AL080133) hypothetical protein [Homo sapiens]	Contains protein domain (PF00435) - Spectrin repeat		56182575, 264259, 60432289, 29331826, 264107, 264905, 264908, 264910, 60170831, 264758, 265010, 265018, 264448, 264288, 264768, 33657109, 264628, 55810784, 18108379, 264634, 56182323, 56526486
1896	85673555 (3791, 3792)		Contains protein domain (PF00627) - UBA domain	UNCLASSIFIED	264907, 265008, 264682, 264686, 21906768, 264629, 264631, 264634, 264555, 264259
1897	80565568 (3793, 3794)	Novel Protein sim. GBank gij728836[sp P39193 ALU6_HUMAN - IIII] ALU SUBFAMILY SP WARNING ENTRY IIII		cadherin	
1898	87617637 (3795, 3796)	Novel Protein sim. GBank gij127560[sp P23249 MV10_MOUSE - PROTEIN MOV-10]		helicase	22278996, 22278998, 22278999, 29331824, 29331825, 60432289, 29331827, 35696052, 29331828, 265008, 265019, 264681, 264682, 264448, 264369, 52644229, 21906765, 21906766, 21906768, 21906769, 60170615, 55810764, 22279000
1899	86673097 (3797, 3798)	Novel Protein sim. GBank gij2909819 (AF031548) - erythrocyte membrane glycoprotein Rh50 [Homo sapiens]	Contains protein domain (PF00909) - Ammonium Transporter Family	glycoprotein	264259, 264508, 264909, 60432229, 264769, 21906765, 21906769
1900	87641858 (3799, 3800)	Novel Protein sim. GBank gij4102881 (AF017250) - vitellogenin precursor [Oreochromis aureus]		UNCLASSIFIED	264683

1801	95196647 (3801, 3802)	Novel Protein sim. GBank gi 585959 sp P38378 S61A_RAT PROTEIN TRANSPORT PROTEIN SEC61 ALPHA SUBUNIT	Contains protein domain (PF00242) - DNA polymerase (viral) N-terminal domain	transport	264488, 52644507, 52645156, 18108396, 52646365, 52646842, 18108397, 56182575, 22278994, 22278995, 56994075, 35696286, 22278997, 22278998, 264490, 60432049, 264259, 29331822, 52645080, 29331824, 29331825, 66714117, 29331826, 60432289, 29331827, 29331828, 35696052, 33656970, 29146498, 264906, 264907, 29331830, 264908, 52644045, 264909, 264112, 265006, 264512, 265008, 264910, 265009, 60170831, 60432229, 60433356, 33657402, 60433438, 55812038, 264758, 33109954, 21908754, 33657084, 52644296, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 18108351, 264448, 264288, 264686, 52644229, 21908765, 21906766, 21906767, 21908769, 55811957, 35695917, 265020, 265021, 52644150, 18108362, 33657023, 264683, 263987, 33657109, 33657182, 27486264, 33657349, 35695763, 18108370, 18108376, 55811576, 35696423, 35695855, 60431850, 264636, 263981, 52644332, 60170394, 83373044, 18108385, 87168518, 60432113, 264564 264107, 263976
1802	80202013 (3803, 3804)	Novel Protein sim. GBank gi 4426613 gb AAD20451  - (AF098796) SLM-1 [Mus musculus]		dna_ma_bind	
1803	87778554 (3805, 3806)	Novel Protein sim. GBank gi 3747107 (AF095741) - unknown [Rattus norvegicus]		UNCLASSIFIED	264259, 29331825, 29331827, 264508, 264807, 265008, 60170831, 60433356, 60433438, 264759, 21906754, 264448, 264288, 265021, 265022, 33657023, 264693, 55811576, 264555, 264556, 22276000
1804	80434213 (3807, 3808)	Novel Protein sim. GBank gi 1352911 sp P47147 VJ80_YEAST - HYPOTHETICAL 80.2 KD PROTEIN IN CPA2-NNF1 INTERGENIC REGION		stuid	264509, 264905, 264806, 264907, 264908, 265007, 264910, 264686, 264768, 264687, 264769, 264693, 264628, 18108374, 264634, 264636, 264637, 264585
1805	95351140 (3809, 3810)	Novel Protein sim. GBank gi 3043714 dbj BA025521  - (AB011167) KIAA0595 protein [Homo sapiens]	Contains protein domain (PF00293) - Bacterial muT protein		264488, 264768, 264769, 264689, 29148629, 35695917, 35696286, 264259, 264692, 18108362, 33657023, 29331824, 33657109, 29146499, 264508, 264509, 264905, 264906, 264907, 66712502, 264908, 264909, 35696423, 35695855, 264510, 264511, 264512, 264910, 264634, 264635, 264637, 264638, 33657402, 264758, 85658542, 264602, 264760, 264761, 264482, 264563, 264762, 264483, 264784, 264566, 264288, 264766
1806	12763822 (3811, 3812)			UNCLASSIFIED	264637

1907	95351144 (3813, 3814)	Novel Protein sim. GBank gij4929585jgb AAD34053.1 AF15181 - (AF151816) CGI-58 protein [Homo sapiens]	Contains protein domain (PF00561) - hydrolase alpha/beta hydrolase fold	65274572, 22278986, 35696286, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 60424269, 29331828, 35696052, 264593, 60433356, 21906754, 55811386, 85658542, 87168559, 265018, 264681, 264682, 264684, 264288, 21906765, 21906766, 21906768, 265020, 265022, 264690, 52644150, 264692, 33657023, 264693, 33657109, 35695855, 264636, 264638, 60432113
1908	95313641 (3815, 3816)	Novel Protein sim. GBank gjj3986770 (AF109906) - NG22 [Mus musculus]	UNCLASSIFIED	264488, 65274572, 56182575, 56181686, 22278995, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 60432289, 29331826, 29331827, 29331828, 28146489, 264905, 264906, 264907, 264908, 66712502, 264909, 56182435, 264510, 264511, 265008, 264910, 60170831, 264592, 60433356, 33657402, 264594, 60433438, 264596, 55812038, 33109954, 52646317, 265011, 265017, 264604, 265018, 265019, 264605, 55811150, 264681, 264448, 264288, 264686, 264688, 264769, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 29148628, 35695917, 265020, 265022, 264691, 264692, 18108364, 65274620, 33657109, 33657349, 35695763, 18108374, 263978, 55810784, 55811576, 35696423, 65274791, 264631, 264632, 264556, 264557, 60170394, 56182323, 83373044, 18108385, 60432113, 22279000, 22279002, 264566, 264488
1909	85514505 (3817, 3818)	Novel Protein sim. GBank gjj224653jgb BAA20813  - (AB002354) KIAA0356 [Homo sapiens]	UNCLASSIFIED	264259, 264508, 264805, 264906, 264907, 264908, 264511, 264910, 264593, 264758, 264764, 264766, 18108370, 264634, 264637, 264488
1910	94216821 (3819, 3820)	Novel Protein sim. GBank gij1351218isp P47226 TES2_MOUSE - TESTIN 2 (TES2) [CONTAINS: TESTIN 1 (TES1)]	Contains protein domain (PF00412) - homeobox LIM domain containing proteins	35696288, 22278996, 22278999, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265007, 264512, 264910, 264758, 265011, 264601, 264602, 264604, 264605, 264761, 264764, 264288, 264766, 264768, 264687, 264769, 35695917, 265021, 52644150, 264692, 264628, 18108370, 264629, 18108372, 18108374, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 18108385, 264565, 264566, 264488

1811	91725345 (3821, 3822)	Novel Protein sim. GBank gi 4809339 gb AAD30184.1 AC006530 - (AC006530) hypothetical protein [Homo sapiens]	Contains protein domain (PF01119) - DNA mismatch repair protein	nuclease	18108394, 56182575, 56182181, 29331826, 29331827, 33656970, 264906, 265007, 264591, 55812038, 87168559, 264448, 264369, 21906765, 21906768, 265022, 264691, 264693, 18108365, 55811576, 264556, 18108385, 18108388
1812	95413518 (3823, 3824)	Novel Protein sim. GBank gi 5689439 dbj BAA83003.1  - (AB028974) KIAA1051 protein [Homo sapiens]	Contains protein domain (PF00098) - Zinc finger, CCHC class	UNCLASSIFIED	22278995, 56994075, 22278996, 22278999, 264259, 29331822, 29331824, 56182181, 29331825, 66714117, 35696052, 264805, 264906, 264907, 264908, 52644045, 56182435, 265007, 265008, 264910, 265009, 264591, 264596, 65274444, 55811386, 87168474, 265011, 87168559, 265018, 265019, 264760, 18108351, 264681, 264369, 264684, 264288, 264686, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 60170615, 264692, 33657023, 264693, 18108376, 55811576, 35696423, 65274791, 264637, 56182323, 83373044, 56526486, 22279002, 264563, 264566
1813	95305546 (3825, 3826)	Novel Protein sim. GBank gi 5032245 ref NP_005665.1 pZNF2 - zinc finger protein (C2H2) homologous to mouse MOK-2			56182575, 22278994, 22278995, 56994075, 22278996, 22278998, 22278999, 29331826, 29331827, 265006, 55812038, 265010, 265017, 265018, 265019, 264681, 18108351, 264683, 264764, 264369, 264288, 264685, 264686, 264769, 21906765, 21906766, 21906768, 21906769, 55811957, 265020, 265022, 264691, 55811576, 264634, 264635, 264638, 56182323, 83373044, 18108385
1814	83423982 (3827, 3828)	Novel Protein sim. GBank gi 4589604 dbj BAA76824.1  - (AB023197) KIAA0980 protein [Homo sapiens]	Contains protein domain (PF00036) - EF hand	struct	56182575, 29331824, 35696052, 264906, 264908, 264828, 264909, 264592, 264758, 87168559, 18108351, 18108354, 264684, 264686, 33657023, 264693, 264628, 264631, 264632, 264634, 264635, 264639
1815	95340459 (3829, 3830)	Novel Protein sim. GBank gi 5689415 dbj BAA82891.1  - (AB028982) KIAA1039 protein [Homo sapiens]		UNCLASSIFIED	264259, 29331824, 29331826, 29331827, 264508, 264909, 265009, 265017, 265019, 264768, 264769, 264689, 264628, 264635, 264637, 264639, 83373044, 264565 264693, 264639
1816	79640761 (3831, 3832)				

1917	87821680 (3833, 3834)	Novel Protein sim. GBank gl 5686991[db][BAA82979.1] - (AB028950) KIAA1027 protein [Homo sapiens]		struct	264769, 264689, 21906765, 21906768, 22278996, 264259, 264691, 264693, 29331824, 29331825, 29331826, 29331828, 264905, 264906, 264828, 264907, 264908, 264909, 264510, 264630, 264910, 264634, 264635, 264636, 264637, 264638, 263981, 264639, 264758, 18108385, 21906754, 265011, 264604, 264583, 18108351, 264762, 264763, 264588, 264764, 264768
1918	95302795 (3835, 3836)	Novel Protein sim. GBank gl 5281517[gb]/AAD41524.1 AF15483 - (AF154831) PV-1 [Rattus norvegicus]		struct	264488, 18108392, 18108357, 21906765, 21906767, 21906768, 56182575, 21906769, 22278994, 35686286, 35685917, 22278996, 22278997, 265021, 265022, 264534, 264690, 264691, 264692, 33657023, 264693, 29331824, 29331825, 33657109, 29331826, 52845129, 35686052, 29331828, 27486262, 27486264, 35695763, 264508, 264905, 264509, 264906, 264628, 264907, 18108370, 264908, 264629, 264909, 18108372, 18108374, 263978, 35686423, 35685855, 264510, 264511, 265006, 265007, 264512, 265008, 264631, 265009, 264910, 264634, 264635, 264555, 264636, 264556, 264637, 264557, 264593, 264638, 264594, 60170394, 264595, 264559, 264596, 83373044, 264758, 52846317, 18108385, 52844298, 56526488, 87188518, 265010, 265011, 87188559, 264600, 264601, 264602, 265017, 264603, 264604, 265018, 264605, 264760, 264761, 264482, 264564, 18108351, 264762, 264682, 264565, 264448, 264764, 264566, 264486, 264567, 264369, 264288, 264766, 264487, 264685
1919	94143847 (3837, 3838)	Novel Protein sim. GBank gl 3878584[emb][CAB01237] - (Z77667) cDNA EST EMBL: C08125 comes from this gene; cDNA EST EMBL: C09753 comes from this gene [Caenorhabditis elegans]		oxidase	22278997, 29331822, 265007, 60170831, 60432229, 60433438, 264448, 264682, 264288, 55811857, 33657023, 33657109, 65274791, 56182323, 22279002
1920	91226953 (3839, 3840)	Novel Protein sim. GBank gl 1809231 (AC000115) - coded for by human cDNAs R76043 (NID:g850725), R65857 (NID:g838495) and H12868 (NID:g877688) [Homo sapiens]		UNCLASSIFIED	264510, 264511, 264512, 264566
1921	79555226 (3841, 3842)	Novel Protein sim. GBank gl 4500997[gb]/AAD24571.1 AF12108 - (AF121081) cAMP inducible 2 protein [Mus musculus]		UNCLASSIFIED	264693



1822	87641863 (3843, 3844)	Novel Protein sim. GBank gi 138595 sp P02845 VIT2_CHICK - VITELLOGENIN II PRECURSOR (MAJOR VITELLOGENIN) [CONTAINS: LIPOVITELIN I (LVI); PHOSVITIN (PV); LIPOVITELIN II (LVI)]; YGP40]		UNCLASSIFIED	264686, 264688, 264490, 18108370, 264909, 18108374, 265008, 264557, 264564, 18108351
1823	84323589 (3845, 3846)	Novel Protein sim. GBank gi 119110 sp P03211 EBN1_EBV - EBNA-1 NUCLEAR PROTEIN		UNCLASSIFIED	264488, 264489, 22278995, 264094, 264259, 35696052, 264509, 264905, 264906, 264907, 264908, 66712502, 264909, 264511, 264512, 265007, 264910, 265009, 264591, 264592, 264593, 264594, 264595, 264758, 264596, 264759, 265011, 265017, 265018, 265019, 55811150, 264681, 264762, 264448, 264764, 264288, 264369, 264766, 264767, 264686, 264687, 264768, 264769, 56181562, 264689, 21908766, 264691, 33657023, 264693, 65274620, 33857109, 18108370, 264628, 264629, 35698423, 264630, 264631, 264632, 264634, 264555, 264636, 264637, 264638, 264639, 264558, 83373044, 87168518, 264563, 264564, 264565, 264566, 264567 29331826, 264908, 264908, 264595, 265017, 265018, 265019, 265021, 264691, 264693, 264637, 18108385, 264565
1824	87338925 (3847, 3848)	Novel Protein sim. GBank gi 3877655 emb CAA96657  - (272511) possible zinc finger protein; cDNA EST EMBL:M89115 comes from this gene; cDNA EST EMBL:D71533 comes from this gene; cDNA EST EMBL:D72314 comes from this gene; cDNA EST EMBL:D75164 comes from this gene; cDNA EST EMBL:C1...	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)		
1825	87628338 (3849, 3850)	Novel Protein sim. GBank gi 4981803 gb AAD36415.1 AE00178 - (AE001788) ribosomal protein S15 [Thermotoga maritima]	Contains protein domain (PF00312) - Ribosomal protein S15	ribosomalprot	22278995, 22278996, 22278997, 264259, 29331824, 66714117, 29146499, 264909, 52644045, 265008, 265009, 264758, 265011, 265017, 264605, 264448, 264288, 264692, 33857109, 18108374, 60170394
1826	88094739 (3851, 3852)	Novel Protein sim. GBank gi 2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma- associated herpesvirus]		struct	264905, 264906, 264907, 264910, 264593, 265018, 264760, 264764, 264288, 264692, 264693, 263978, 264631, 264634, 264637, 264563 264689, 264631
1827	85654857 (3853, 3854)	Novel Protein sim. GBank gi 3043632 dbj BAA25480  - (AB011128) KIAA0554 protein [Homo sapiens]			
1828	87799054 (3855, 3856)	Novel Protein sim. GBank gi 1665761 dbj BAA13377  - (D87433) KIAA0246 [Homo sapiens]	Contains protein domain (PF00193) - Extracellular link domain	Inf	264488, 264259, 265017, 265021, 264692
1829	86997236 (3857, 3858)	Novel Protein sim. GBank gi 5001693 gb AAD37247.1 AF13432 - (AF134321) chimeric AFGP/hypophosphagen-like serine protease precursor [Dissostichus mawsoni]		UNCLASSIFIED	264508, 264591, 33657402, 265017, 264768, 264632, 264556, 264639

1830	87889128 (3859, 3860)	Novel Protein sim. GBank gi 1709230 sp P52963 NBL4_MOUSE - NBL4 PROTEIN		phosphatase	35696286, 29331828, 264905, 264907, 264908, 264909, 264511, 264910, 264758, 264601, 265017, 265019, 264805, 264780, 264764, 264766, 264688, 264769, 265022, 35696423, 264638, 60432113
1831	87797278 (3861, 3862)	Novel Protein sim. GBank gi 404634 (U01840) - serine/threonine kinase [Mus musculus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264906, 264908, 60432229, 264758, 264764, 264288, 265020, 264692, 264634, 264637, 264684, 264691, 264635
1832	15030972 (3863, 3864)			UNCLASSIFIED	264595
1833	11613668 (3865, 3866)			struct	56162575, 56182435, 264510, 264757, 264758, 55812038, 55811388, 265018, 55811150, 21906765, 264681, 264631, 264635, 264637
1834	84426360 (3867, 3868)	Novel Protein sim. GBank gi 115748 db BAA36494  - (AB022023) nonmuscle myosin heavy chain B [Bos taurus]		UNCLASSIFIED	264886, 265011, 264511, 264905, 18108351, 264564, 264681, 264259, 18108370, 264566, 264764, 264369, 264595
1835	87752511 (3869, 3870)			UNCLASSIFIED	60432289, 265007, 265010, 265011, 265019, 33657109, 18108374
1836	95414338 (3871, 3872)	Novel Protein sim. GBank gi 4827040 ref NP_005110.1 pTRAP - thyroid hormone receptor-associated protein, 150 kDa subunit			
1837	94847141 (3873, 3874)	Novel Protein sim. GBank gi 543187 pir J S3771 - ankyrin, erythrocyte - mouse	Contains protein domain (PF00023) - Ank repeat	kinase	35658542, 21906767, 35695917, 60170815, 264693, 33857109
1838	87403277 (3875, 3876)	Novel Protein sim. GBank gi 4544431 gb AA022340.1 AC00695 - (AC006955) hypothetical protein [Arabidopsis thaliana]	Contains protein domain (PF00888) - Cullin family	collagen	264488, 29146488, 264905, 264559
1839	91004978 (3877, 3878)	Novel Protein sim. GBank gi 500858 db BAA03210  - (D14166) 50kDa lectin [Bombyx mori]		UNCLASSIFIED	65274572, 22278899, 60432289, 56182435, 60433356, 265017, 21906765, 21906766, 21906788, 55811957, 27486284, 35696423, 60432113, 264564
1840	87348810 (3879, 3880)	Novel Protein sim. GBank gi 1946300 emb CAA73132  - [Y12529] hypothetical protein [Silene latifolia]	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	264488, 29331822, 264448, 264683, 264288, 265020, 33857023, 264631
1841	94147177 (3881, 3882)	Novel Protein sim. GBank gi 4206386 (AF060570) - rig-1 protein [Mus musculus]		UNCLASSIFIED	56994075, 22278998, 264259, 29331824, 29331827, 264905, 265008, 33857084, 265017, 265018, 264288, 264687, 21906765, 21906766, 21906767, 265020, 52644150, 27486284, 83373044, 18108387, 60432113, 22279002, 264565
1842	87641870 (3883, 3884)	Novel Protein sim. GBank gi 4927204 gb AAD33049.1 AF13391 - (AF133911) ARL-6 interacting protein-4 [Mus musculus]		UNCLASSIFIED	264488, 18108398, 29331825, 27486261, 264509, 18108370, 18108374, 264482
1843	94325288 (3885, 3886)	Novel Protein sim. GBank gi 3122952 sp O15736 TIPD_DICD1 - TIPD PROTEIN	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	22278998, 29331822, 29331827, 35696052, 264511, 265009, 264592, 60432228, 265017, 265018, 265019, 264684, 264692, 33657109, 65274791, 264636

1944	94232958 (3887, 3888)	Novel Protein sim. GBank gij1799570[dbj][BAA13432] - (D87671) TIP120 [Rattus norvegicus]		UNCLASSIFIED	65274572, 22278994, 22278995, 22278996, 22278998, 22278999, 264259, 52645080, 29331824, 29331826, 29331827, 66712502, 56182435, 264512, 264910, 60170831, 60432229, 60433356, 33657402, 60433438, 264596, 33109954, 21906754, 87168474, 87168559, 265017, 265018, 265019, 18108351, 264369, 264686, 264768, 21906765, 21906766, 21906767, 21906769, 35695917, 265020, 265021, 60170515, 264692, 33657023, 18108370, 18108374, 35696423, 35695855, 264634, 60170394, 264639, 83373044, 18108385, 56526486, 87168518, 60432113, 22279000, 264563
1945	87641872 (3889, 3890)	Novel Protein sim. GBank gij4927204[gbj][AAD33049.1][AF13391 - (AF133911) ARL-6 interacting protein-4 [Mus musculus]		UNCLASSIFIED	264488, 22278996, 264510, 264511, 18108351, 264683, 264486, 264567
1946	87443990 (3891, 3892)	Novel Protein sim. GBank gij2498104[spj][Q27969][AD50_BOVIN - ADRENAL MEDULLA 50 KD PROTEIN]		UNCLASSIFIED	60432289, 29331827, 35696052, 265007, 265008, 60433356, 60433438, 264369, 56181562, 21906767, 52644150, 284693, 27486264, 264637, 87168518, 264563
1947	86438862 (3893, 3894)	Novel Protein sim. GBank gij3914801[spj][O54886][RPA2_RAT - DNA-DIRECTED RNA POLYMERASE I 135 KD POLYPEPTIDE (RNA POLYMERASE I SUBUNIT 2) (RPA135) (RNA POLYMERASE I 127 KD SUBUNIT)]		mapolymerase	22278998, 264905, 264906, 264908, 264909, 264512, 264758, 264762, 264682, 264683, 284764, 264288, 264768, 264688, 21906768, 264693, 18108374, 35695855, 264635, 264637, 264638, 18108385, 22279002
1948	85199174 (3895, 3896)	Novel Protein sim. GBank gij5420387[embj][CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		struct	264909, 60170831, 284591, 264594, 235010, 265011, 264764, 264369, 264689, 264631, 264638
1949	7640129 (3897, 3898)	Novel Protein sim. GBank gij3876766[embj][CAA93466.1] - (Z69637) predicted using GeneFINDER; Similarity to E.coli hypothetical protein YCAC (SW:YCAC_ECOLI) [Caenorhabditis elegans]	Contains protein domain (PF00857) - Isochorismalase family	UNCLASSIFIED	264369
1950	87788531 (3899, 3900)	Novel Protein sim. GBank gij2626753[dbj][BAA23424] - (AB008782) sulfate transporter [Arabidopsis thaliana]	Contains protein domain (PF00916) - Sulfate transporter family	transport	264488, 264768, 264688, 264689, 264259, 29331822, 33657109, 29331828, 264103, 264509, 18108370, 35695855, 264510, 265008, 265009, 33657402, 18108385, 265018, 264563, 55811150, 18108351, 264369, 264288, 18108354
1951	86988253 (3901, 3902)	Novel Protein sim. GBank gij4929633[gbj][AAD34077.1][AF15184 - (AF151840) CGI-82 protein [Homo sapiens]	Contains protein domain (PF00106) - short chain dehydrogenase	reductase	56182575, 22278997, 52645080, 29331824, 29331825, 29331827, 55812038, 52646317, 265018, 265019, 264369, 21906765, 21906767, 55811957, 265020, 265021, 33657023, 264693, 35695763, 56182323, 22279002
1952	87069775 (3903, 3904)	Novel Protein sim. GBank gij4929633[gbj][AAD34077.1][AF15184 - (AF151840) CGI-82 protein [Homo sapiens]	Contains protein domain (PF00106) - short chain dehydrogenase	reductase	264591, 264593, 264594, 264595, 264555, 264556, 264557, 264558, 264565

1853	20470371 (3905, 3906)	Novel Protein sim. GBank gij1168715 sp P31721 C1QB_RAT - COMPLEMENT C1Q SUBCOMPONENT, B CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264259, 264558
1854	91226025 (3907, 3908)	Novel Protein sim. GBank gij4240271 dbj BAA74914.1  - (AB020698) KIAA0891 protein [Homo sapiens]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	264488, 263994, 18108394, 35696286, 22278998, 29331822, 66714117, 29331826, 29331827, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265008, 265007, 265008, 264910, 265009, 60170831, 60432229, 265011, 265017, 264603, 264604, 265019, 18108351, 264685, 264768, 264769, 35895917, 265020, 265021, 264691, 33657023, 264692, 33657109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264635, 264636, 264555, 264556, 264638, 264557, 264639, 264558, 264559, 83373044, 18108385, 264563, 264564, 264568, 264488, 264567
1855	95308310 (3909, 3910)			UNCLASSIFIED	56182575, 56181686, 22278994, 22278999, 264259, 29331822, 56182181, 29331824, 29331825, 29331826, 29331827, 35696052, 264508, 29331830, 265008, 265009, 264591, 55812038, 87168474, 265017, 265018, 265019, 264448, 264768, 21906765, 21906766, 21806767, 55811957, 265020, 265021, 52645129, 33657109, 27486264, 33657349, 35695763, 60431528, 18108374, 55811576, 35695855, 264635, 60431850, 264639, 83373044
1856	95092121 (3911, 3912)	Novel Protein sim. GBank gij1665821 dbj BAA134071 - (D87469) Similar to D.melanogaster cadherin-related tumor suppressor [Homo sapiens]	Contains protein domain (PF00029) - Cadherin domain	cadherin	264488, 56182575, 22278996, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 264508, 264906, 264908, 264909, 264511, 265007, 264910, 264591, 55812038, 265010, 265018, 18108351, 264768, 56181562, 264689, 21906768, 21906769, 265022, 264691, 264628, 18108374, 55810764, 55811576, 35695855, 264631, 264632, 264635, 264637, 264639, 60170394, 56182323, 83373044, 18108385, 22279000, 22279002, 264563, 264564

1957	94326510 (3913, 3914)	Novel Protein sim. GBank gij4589674[dbj BAA76856.1] - (AB023229) KIAA1012 protein [Homo sapiens]		UNCLASSIFIED	52648842, 56182575, 22278997, 22278998, 22278999, 29331824, 66714117, 29331827, 29146498, 284593, 33657402, 33109954, 87168474, 265018, 264448, 264369, 264288, 264766, 21906766, 21908767, 21906768, 21906769, 265020, 265021, 264692, 65274620, 27486264, 33657349, 27486265, 35695855, 22279002, 264482
1958	95313902 (3915, 3916)	Novel Protein sim. GBank gij4240227[dbj BAA74892.1] - (AB020676) KIAA0869 protein [Homo sapiens]		UNCLASSIFIED	22278998, 264092, 264094, 264259, 60432049, 29331824, 56182181, 66714117, 264107, 264109, 264909, 264511, 60170831, 60432229, 21906754, 265010, 21906769, 35695917, 265022, 65274620, 263967, 263976, 35686423, 264631, 264632, 264634, 264635, 18108385, 22279000, 22279002, 264593, 265019
1959	85701470 (3917, 3918)	Novel Protein sim. GBank gij2281983[emb CAB10860] - (Z98056) hypothetical protein [Schizosaccharomyces pombe]		ubiquitin	
1960	80308608 (3919, 3920)	Novel Protein sim. GBank gij2274851[dbj BAA21515] - (D64159) 3-7 gene product [Homo sapiens]		struct	264905, 264906, 264907, 264908, 264909, 265006, 265007, 264910, 264595, 265017, 264604, 265018, 18108351, 264764, 264369, 264766, 264768, 21906765, 18108368, 264629, 18108379, 264635, 264636, 264637, 264638, 264486
1961	16292607 (3921, 3922)				264635
1962	91003385 (3923, 3924)			UNCLASSIFIED	65274572, 264592, 264593, 265019, 264691
1963	80936017 (3925, 3926)	Novel Protein sim. GBank gij3721653[dbj BAA3358.1] - (AB012933) acyl-CoA synthetase 5 [Rattus norvegicus]		eph	65274572, 18108398, 35686286, 29331825, 60432289, 29331827, 264828, 265006, 265009, 60433356, 60433438, 21906754, 265020, 265021, 33657023, 33657109, 27486265, 35695855, 264555
1964	94317605 (3927, 3928)	Novel Protein sim. GBank gij5262638[emb CAB45757.1] - (AL080169) hypothetical protein [Homo sapiens]		cadherin	264488, 264092, 264259, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265007, 265009, 264910, 264592, 264593, 264594, 264595, 264758, 264600, 264603, 264604, 264605, 264760, 264762, 264448, 264764, 264288, 264685, 264766, 264768, 264769, 21906766, 264691, 264692, 264693, 18108370, 264628, 264629, 18108374, 264630, 264631, 264634, 264636, 264637, 264638, 18108382, 83373044, 18108385, 264483, 264564, 264565, 264566, 264486, 264567
1965	94317445 (3929, 3930)	Novel Protein sim. GBank gij4107017[dbj BAA36294] - (AB001773) PEM-6 [Ciona savignyi]	Contains protein domain (PF01428) - AN1-like Zinc finger	ubiquitin	264488, 264510, 264760, 264768, 264486

1866	94192058 (3931, 3932)	Novel Protein sim. GBank gi4929707 gb AAD34114.1 AF15187 - (AF151877) CGI-119 protein [Homo sapiens]	Contains protein domain (PF01027) - Uncharacterized protein family	glycoprotein	22278999, 264092, 264259, 29331826, 29331828, 29146498, 264595, 265011, 264448, 18108354, 264288, 264684, 264766, 264885, 264686, 265022, 264891, 264692, 18108370, 18108377, 264555, 18108381, 18108385, 264486, 264567
1867	87396123 (3933, 3934)	Novel Protein sim. GBank gi2957270 (AF044576) - phospholipase C PLC210 [Caenorhabditis elegans]	Contains protein domain (PF00368) - Phosphatidylinositol-specific phospholipase C, X domain	esterase	29331824, 265010, 265017, 264288, 21906784, 263981, 56526486
1868	88095641 (3935, 3936)	Novel Protein sim. GBank gi2564953 (AF030001) - unknown [Mus musculus]	Contains protein domain (PF00008) - EGF-like domain	oncogene	35696286, 264905, 264509, 264906, 264907, 264908, 264908, 264511, 264512, 265008, 264910, 265009, 264594, 264757, 264758, 264604, 264605, 264760, 264762, 264682, 264764, 264685, 264768, 264767, 264689, 264691, 264693, 264628, 264629, 35698423, 35695855, 264631, 264632, 264634, 264635, 264636, 264637, 18108380, 264564, 264565, 264566, 264567
1869	84328529 (3937, 3938)	Novel Protein sim. GBank gi2911274 (U20329) - spidroin 1 [Nephila clavipes]		UNCLASSIFIED	22278995, 22278996, 35696052, 264906, 264908, 18108351, 264482
1870	80596049 (3939, 3940)	Novel Protein sim. GBank gi4050087 (AF109907) - S164 [Homo sapiens]		UNCLASSIFIED	264908, 264288, 264766, 264636
1871	94843914 (3941, 3942)	Novel Protein sim. GBank gi134206 sp P09593 SANT_PLAFV - S-ANTIGEN PROTEIN PRECURSOR		collagen	264488, 264489, 22278998, 264259, 60432048, 66714117, 29331826, 60432289, 29331827, 35696052, 264508, 264905, 264508, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264910, 264591, 264592, 60432229, 60433356, 264595, 264596, 264600, 264604, 264605, 264760, 18108351, 264448, 264764, 264288, 264766, 264768, 264769, 21906765, 33657023, 264692, 18108370, 264629, 35696423, 65274791, 35695855, 264632, 264635, 264555, 264636, 264637, 264638, 264639, 18108385, 60432113, 22279000, 264563, 264564, 264565, 264566, 264486
1872	87645444 (3943, 3944)	Novel Protein sim. GBank gi4519623 dbj BAA75671.1] - (AB017616) homologous to the yeast YGR163 gene [Mus musculus]	Contains protein domain (PF01462) - Leucine rich repeat N-terminal domain		22278998, 264259, 29331822, 56182181, 60432289, 29331827, 52644045, 264909, 265006, 264511, 265008, 52644296, 265018, 265019, 264761, 264689, 21906768, 21906769, 264691, 264693, 33657109, 33657182, 264556, 52644332, 264558, 60432113
1873	86395533 (3945, 3946)			UNCLASSIFIED	29331826, 264692, 35696423, 264631, 264555, 264556, 264557, 264558, 264559
1874	80396629 (3947, 3948)	Novel Protein sim. GBank gi3309543 (AF036382) - MLL [Fugu rubripes]		UNCLASSIFIED	264682, 264764, 264563

1975	94316479 (3949, 3950)			UNCLASSIFIED	264488, 66714117, 29331826, 29331828, 56182435, 265006, 264757, 55812038, 265010, 265017, 264369, 55811957, 65274781, 35695855, 56182323, 60432113, 264259, 35696052, 265018, 265020, 265021, 33657109, 56526486
1976	94358814 (3951, 3952)			UNCLASSIFIED	264908, 264596, 265021, 264566
1977	94852664 (3953, 3954)	Novel Protein sim. GBank gi 2499526 sp Q07782 NASU_RAT - SODIUM/SULFATE COTRANSPORTER (NA(+)/SULFATE COTRANSPORTER)		homeobox	
1978	87447845 (3955, 3956)	Novel Protein sim. GBank gi 103421 pir A33471 - transcription factor NTF1 - fruit fly (Drosophila melanogaster) (fragment)		transcription factor	60170831, 264566
1979	87627708 (3957, 3958)	Novel Protein sim. GBank gi 2244815 emb CAB10238.1  - (Z97336) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	29331826, 29146498, 264905, 264907, 265007, 265009, 265010, 265018, 264686, 18108359, 21906768, 35695917, 265020, 60170615, 264693, 18108368, 18108370, 264631, 264635, 264556, 264558, 18108384, 22279000, 264565
1980	86577059 (3959, 3960)	Novel Protein sim. GBank gi 4759290 ref NP_004642.1 pUSP1 - Ubiquitin carboxyl-terminal hydrolase, X-linked		ubiquitin	264489
1981	87606974 (3961, 3962)	Novel Protein sim. GBank gi 4406693 gb AA020060  - (AF131849) Unknown [Homo sapiens]		UNCLASSIFIED	22278995, 22278997, 22278999, 264259, 29331825, 29331828, 29146498, 29146499, 264107, 264906, 264910, 264595, 21906754, 265010, 265017, 265018, 265019, 264448, 264288, 21906767, 33657023, 27486264, 18108370, 18108374, 18108376, 264630, 264631, 264635, 18108385, 87168518, 22279000, 264482, 264564
1982	90995367 (3963, 3964)	Novel Protein sim. GBank gi 5689523 dbj BAA83045.1  - (AB029016) KIAA1093 protein [Homo sapiens]			65274572, 29331825, 35696052, 33656870, 264909, 265008, 55811386, 264760, 264686, 264691, 27486264
1983	95098668 (3965, 3966)	Novel Protein sim. GBank gi 3417297 (AC002310) - Unknown gene product [Homo sapiens]	Contains protein domain (PF000096) - Zinc finger, C2H2 type	transcription factor	22278996, 35696286, 264259, 29331822, 20281099, 29331824, 60432289, 29331827, 264509, 264905, 264906, 264907, 66712502, 264908, 52844045, 264909, 264510, 264512, 264910, 265009, 264591, 264592, 60433356, 60433438, 264758, 265010, 264600, 264603, 264604, 264760, 264762, 264763, 264784, 264766, 264687, 264768, 264769, 21906765, 55811957, 35695917, 264690, 264692, 264693, 264628, 264629, 263978, 18108379, 35696423, 35695855, 20281071, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 83373044, 18108385, 22279000, 22279002

1884	85760889 (3867, 3968)	Novel Protein sim. GBank gij2896895[emb]CAA17174.1] - (AL021897) tadD14 [Mycobacterium tuberculosis]		synthase	264688, 21908768, 55811957, 56994075, 265020, 265021, 22278999, 265022, 264259, 29331822, 33657182, 29146499, 264628, 18108370, 264908, 264629, 55811576, 35695855, 265006, 265007, 264591, 21906754, 33657084, 265010, 265017, 265019, 264288
1885	85636897 (3969, 3970)	Novel Protein sim. GBank gij5712131[gb]AAD47379.1[AF12049] - (AF120499) DEM1 protein [Homo sapiens]		glycoprotein	264760, 264288, 263978, 55811576, 264637, 56182323, 18108385, 264564
1886	80200507 (3971, 3972)			UNCLASSIFIED	264488, 264629
1887	87011117 (3973, 3974)	Novel Protein sim. GBank gij4868443[gb]AAD31319.1[AF14457] - (AF144573) Mx-Interacting protein kinase PKM [Mesocricetus auratus]	Contains protein domain (PF000069) - Eukaryotic protein kinase domain		22278999, 29331830, 265007, 265018, 21906768, 33657023, 264692, 264693, 18108377, 264635, 60170394, 22279002
1888	94122108 (3975, 3976)			UNCLASSIFIED	264905, 264906, 264907, 264908, 264909, 264910, 264591, 264593, 264758, 264764, 264686, 264768, 265021, 264692, 264628, 264629, 35695855, 264630, 264635, 264636, 264637, 264638, 264639, 264483
1889	91225225 (3977, 3978)	Novel Protein sim. GBank gij2801701 (AF042379) - spindle pole body protein spc97 homolog GCP2 [Homo sapiens]		tubulin	60432049, 60432289, 52644045, 56182435, 264112, 265007, 33657402, 52644229, 21906765, 21906768, 21906769, 55811957, 33657023, 263967, 33657109, 18108370, 22279000, 22279002
1890	85698888 (3979, 3980)	Novel Protein sim. GBank gij5701727[db]BAA83074.1] - (AB024728) alpha-1,3-D-mannoside beta-1,4-N-acetylglucosaminyltransferase IV-homologue [Homo sapiens]			264508, 264757, 264764, 18108381
1891	95353114 (3981, 3982)	Novel Protein sim. GBank gij4240287[db]BAA74922.1] - (AB020706) KIAA0899 protein [Homo sapiens]	Contains protein domain (PF01602) - Adaplin N terminal region	glycoprotein	18108394, 56182575, 22278994, 35696286, 56994075, 22278997, 22278999, 29331822, 29331824, 29331825, 60432289, 29331828, 264508, 264906, 264907, 264908, 56182435, 264510, 265007, 21906754, 33109954, 87168474, 265017, 265018, 265019, 264762, 18108351, 264763, 264683, 264369, 264288, 264685, 264766, 264687, 264769, 21906765, 21906768, 21906769, 55811957, 265020, 60431528, 263974, 18108379, 35695855, 264555, 264557, 264639, 83373044, 18108384, 87168518, 60432113, 22279000, 22279002, 264564, 264486
1892	95317232 (3983, 3984)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]	Contains protein domain (PF000096) - Zinc finger, C2H2 type	struct	28331827, 264906, 264907, 264909, 265007, 264603, 264766, 264686, 264768, 21906768, 264628, 264635, 264636, 18108385, 56526486, 264566, 264567
1893	80054763 (3985, 3986)	Novel Protein sim. GBank gij2565091 (U80761) - CTG26 alternate open reading frame [Homo sapiens]		UNCLASSIFIED	264592, 35696423



1994	84325114 (3987, 3988)	Novel Protein sim. GBank gi 5630077 gb AAD45822.1 AC006017 - (AC006017) similar to ALR; similar to AAC51735 (PID:g2358287) [Homo sapiens]	Contains protein domain (PF000856) - SET domain	mapolymerase	264488, 22278997, 22278998, 22278999, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264907, 264908, 265008, 60432229, 33657402, 60433356, 264757, 60433438, 264758, 33109954, 265011, 265017, 265018, 265019, 264684, 264369, 264685, 264686, 264768, 21906765, 21906767, 21906768, 21906769, 265020, 264690, 18108362, 264693, 65274620, 18108370, 264635, 264555, 264556, 264557, 56182323, 83373044, 56526486, 22279000, 22279002, 264564
1995	85414353 (3989, 3990)	Novel Protein sim. GBank gi 4827040 ref NP_005110.1 pTRAP - thyroid hormone receptor-associated protein, 150 kDa subunit	UNCLASSIFIED	UNCLASSIFIED	264488, 18108396, 22278994, 56994075, 22278998, 35896286, 22278997, 22278999, 264259, 28147620, 56182181, 29331824, 60432289, 29331826, 29331827, 35696052, 29146499, 264905, 264907, 66712502, 56182435, 265008, 265007, 265008, 265009, 60431735, 60433356, 33657402, 264595, 55812038, 33657084, 55811386, 85658542, 265010, 265011, 265017, 265018, 265019, 264761, 264762, 264448, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 56181562, 264689, 21906765, 21906766, 21906767, 29148627, 21906768, 21906769, 29148629, 29148784, 265020, 265021, 264690, 18108361, 264693, 27486262, 27486264, 27486265, 18108370, 60431528, 18108374, 18108377, 35698423, 55811578, 65274791, 35695855, 264631, 264634, 264635, 264555, 264636, 60431850, 264557, 264558, 264559, 83373044, 20798451, 87188518, 264404, 60432113, 264567, 264564
1996	80254186 (3991, 3992)	Novel Protein sim. GBank gi 791146 emb CAA60020  - (X86028) extensin-like protein [Vigna unguiculata]		UNCLASSIFIED	
1997	87028423 (3993, 3994)	Novel Protein sim. GBank gi 2642034 (AF034547) - protein phosphatase M130 myosin binding subunit [Ovis aries]	Contains protein domain (PF00023) - Ank repeat	phosphatase	264908, 264909, 264592, 264593
1998	85262704 (3995, 3996)	Novel Protein sim. GBank gi 4589634 dbj BAA76839.1  - (AB023212) KIAA0995 protein [Homo sapiens]		UNCLASSIFIED	264113, 264685, 264555, 264567

1999	94324903 (3997, 3998)	Novel Protein sim. GBank gij5225312igbAAD40846.1(AF072441) calchneurin binding protein cabin 1 [Homo sapiens]	Contains protein domain (PF00515) - TPR Domain	UNCLASSIFIED	18108394, 18108397, 35696286, 60424269, 29331827, 29331828, 35696052, 265006, 264512, 55811386, 265010, 265018, 265019, 55811150, 18108351, 264763, 264682, 264369, 264685, 264686, 56181562, 265020, 264691, 33657023, 264693, 33657108, 27486264, 18108370, 18108379, 35695855, 264634, 264635, 264636, 264555, 264557, 56182323, 18108382, 264559, 83373044, 60432113, 22279000, 264563, 264564, 264568
2000	95413705 (3999, 4000)	Novel Protein sim. GBank gij1723232[sp]Q10155[YATA_SCHPO - HYPOTHETICAL 90.6 KD PROTEIN CTD4.10 IN CHROMOSOME 1		UNCLASSIFIED	52646365, 52646842, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 29331826, 29331827, 29331828, 35696052, 264106, 29331830, 52644045, 265007, 265008, 60170831, 264592, 264593, 33657402, 60433438, 21906754, 52644296, 265017, 265018, 265019, 264761, 264389, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 60170815, 52644150, 33657023, 65274620, 52645129, 27486261, 27486262, 27486264, 35695763, 35696423, 35695855, 264631, 52644332, 56182323, 60170394, 83373044, 56526486, 22278002, 264566, 264567
2001	95072534 (4001, 4002)	Novel Protein sim. GBank gij107560[pir][B38637 - Ras inhibitor (clone JC265) - human (fragment)		UNCLASSIFIED	264769, 52644229, 65274572, 21906768, 22278996, 35696286, 35695917, 265020, 22278999, 264534, 264490, 264259, 264692, 60432289, 33657109, 35696052, 264508, 264509, 18108370, 60431528, 18108374, 35696423, 65274791, 35695855, 264510, 264511, 264512, 265009, 264634, 264636, 264555, 264556, 264638, 264557, 264558, 264559, 60433438, 83373044, 264759, 18108385, 265011, 264600, 264601, 60432113, 264603, 264604, 264605, 264448, 264288, 264765
2002	80236368 (4003, 4004)	Novel Protein sim. GBank gij729433[sp]P38657[ER60_BOVIN - PROBABLE PROTEIN DISULFIDE ISOMERASE ER-60 PRECURSOR (ERP60) (58 KD MICROSOMAL PROTEIN) (P58) Novel Protein sim. GBank gij86388[pir][A27040 - neurofilament triplet M protein - chicken (fragment)]	Contains protein domain (PF00085) - Thioredoxin	- isomerase	264907, 265006, 264910, 264603, 264692, 264629, 18108374, 264556, 264557
2003	80074449 (4005, 4006)			UNCLASSIFIED	264905, 264906, 264908, 264910, 264596, 265017, 18108351, 264692, 264629, 264634, 264565

2004	95317318 (4007, 4008)	Novel Protein sim. GBank gi 4884249 emb CAB3230.1  - (AL049996) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	52645156, 52646842, 52646385, 56182575, 22278994, 22278995, 56994075, 22278996, 35696286, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 29331824, 29331826, 29331827, 35696052, 29331828, 33656970, 29331830, 264908, 264592, 60433356, 33657402, 52646317, 21906754, 33657084, 52644296, 87168474, 87168559, 265017, 265018, 265019, 264763, 264683, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 33657349, 27486265, 35695763, 18108374, 18108376, 18108377, 35696423, 35695855, 264631, 52644332, 264558, 18108385, 56526486, 87168518, 60432113, 264483, 264488, 264906
2005	87400864 (4008, 4010)	Novel Protein sim. GBank gi 3878501 emb CAA87795  - (Z47812) similar to ubiquitin carboxyl-terminal hydrolase; cDNA EST EMBL:D33366 comes from this gene; cDNA EST EMBL:D33965 comes from this gene; cDNA EST EMBL:D33822 comes from this gene; cDNA EST EMBL:D34547 comes from this ge...		ubiquitin	
2006	95351177 (4011, 4012)	Novel Protein sim. GBank gi 4106673 emb CAA22613  - (AL035064) queuine tRNA-ribosyltransferase [Schizosaccharomyces pombe]	Contains protein domain (PF01702) - Queuine tRNA-ribosyltransferase	UNCLASSIFIED	56182575, 56181686, 22278995, 22278996, 22278997, 22278998, 22278999, 264259, 56182181, 60424268, 66714117, 35696052, 264906, 66712502, 264909, 264510, 60433356, 85558542, 265010, 265018, 265019, 264882, 264448, 264288, 264768, 29148627, 21906769, 29148784, 35695917, 60170615, 264691, 33657023, 65274620, 33657109, 55810764, 55811576, 35695855, 87168518, 60432113, 264563, 264482
2007	94325556 (4013, 4014)	Novel Protein sim. GBank gi 2662181 dbj BAA23712  - (AB007900) HH0452 cDNA clone for KIAA0440 has a 438-bp insertion at position 1711 of the sequence of KIAA0440. [Homo sapiens]		UNCLASSIFIED	264488, 263994, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 265007, 264910, 264592, 264595, 264758, 265011, 264760, 284762, 264764, 264768, 264685, 264767, 264768, 264769, 55811957, 35695917, 265020, 264691, 264693, 264628, 264629, 65274791, 35695855, 264631, 264632, 264634, 264635, 264637, 264638, 264639, 264566
2008	85084428 (4015, 4016)	Novel Protein sim. GBank gi 1550763 emb CAA69257  - (Y07960) homeodomain protein [Mus musculus]	Contains protein domain (PF00046) - Homeobox domain	homeobox	264909, 264768, 35695855

2009	85749240 (4017, 4018)	Novel Protein sim. GBank gi 3882305 db BAA34512.1  - (AB018335) KIAA0792 protein [Homo sapiens]	UNCLASSIFIED	22278999, 264259, 264910, 264591, 265017, 264681, 264683, 21908768, 264691, 33657182, 33657349, 264631, 87168518, 264404, 22279002, 264563
2010	95422458 (4019, 4020)	Novel Protein sim. GBank gi 5262629 emb CAB45753.1  - (AL080164) hypothetical protein [Homo sapiens]	eph Contains protein domain (PF00057) - Low-density lipoprotein receptor domain class A	52644507, 52645156, 52646365, 52646842, 18108397, 65274572, 22278994, 58994075, 35896286, 22278996, 22278997, 22278999, 264258, 29331822, 52645080, 29331824, 28331825, 28331826, 28331827, 28331828, 264511, 265007, 264512, 265008, 265009, 60432229, 60433356, 21906754, 52646317, 33108954, 52644296, 87168474, 87168559, 265017, 265018, 265019, 264681, 264685, 264687, 52644229, 264689, 21906765, 21906766, 21906767, 21906768, 35895917, 265020, 52644150, 264691, 264692, 33657023, 263987, 52645129, 35695763, 18108376, 35696423, 65274791, 35695855, 264631, 264634, 60431850, 264637, 264638, 52644332, 60170394, 18108385, 87168518, 22279002, 264584, 264585, 264586, 264587, 56182575, 56994075, 22278999, 264259, 29331824, 29331826, 29331827, 29331828, 35896052, 264906, 66712502, 265006, 265007, 265008, 265010, 265011, 265017, 265019, 264681, 264448, 264683, 264369, 264288, 264685, 264766, 264687, 21906765, 21906767, 21906768, 21906769, 265020, 265022, 264691, 33657023, 65274620, 33657109, 264629, 264557, 264559, 83373044, 87168518, 60432113, 22279002
2011	94328149 (4021, 4022)	Novel Protein sim. GBank gi 3347953 (AF076183) - cytosolic sorting protein PACS-1a [Rattus norvegicus]	UNCLASSIFIED	264259, 29331822, 29331824, 29331825, 29146498, 87168559, 265019, 264682, 264288, 264686, 21906764, 265020, 265022, 33657023, 264693, 33657109, 55811576, 264632, 264558, 56182323, 264639, 18108383, 18108384, 18108388, 22279000, 22279002, 264587
2012	87772137 (4023, 4024)	Novel Protein sim. GBank gi 1086678 (U41020) - coded for by C. elegans cDNA yk100g4.5; coded for by C. elegans cDNA yk100g4.3; weakly similar to human SREBP-2 basic-helix-loop-helix-leucine zipper transcription factor [Caenorhabditis elegans]	UNCLASSIFIED	264259, 29331822, 29331824, 29331825, 29146498, 87168559, 265019, 264682, 264288, 264686, 21906764, 265020, 265022, 33657023, 264693, 33657109, 55811576, 264632, 264558, 56182323, 264639, 18108383, 18108384, 18108388, 22279000, 22279002, 264587
2013	94843842 (4025, 4026)	Novel Protein sim. GBank gi 4507985 ref NP_003427.1 pZNF1 - zinc finger protein 135 (clone pHZ-17)	dna_ma_bind Contains protein domain (PF00096) - Zinc finger, C2H2 type	18108398, 264908, 265007, 265010, 265018, 265019, 264689, 21906767, 265020, 264692
2014	87347840 (4027, 4028)	Novel Protein sim. GBank gi 12720 sp P20538 MYP0_HETFR - MYELIN P0 PROTEIN PRECURSOR	UNCLASSIFIED	264488, 29331826, 264907, 264636, 264555, 264639, 264558

2015	88094922 (4029, 4030)	Novel Protein sim. GBank gij181286[pri][S22697 - extensin - Volvox carteri (fragment)]		UNCLASSIFIED	56182575, 35696286, 264259, 35696052, 264508, 264906, 264907, 264510, 264512, 87168474, 265010, 264681, 264288, 264689, 264628, 35696423, 35695855, 264639, 264563, 264564
2016	85298641 (4031, 4032)	Novel Protein sim. GBank gij285046[pri][S26413 - t-complex protein Tc-10 - mouse]		struc1	264102, 264508, 264110, 265009, 33109954, 21906768, 265021, 33657109, 27486262, 263972, 18108374, 263976, 264555, 264564
2017	79464293 (4033, 4034)			UNCLASSIFIED	264685, 264636
2018	79637067 (4035, 4036)	Novel Protein sim. GBank gij124735[pri][S18175][NVO_PIG - INVOLUCRIN]			264693
2019	87787900 (4037, 4038)	Novel Protein sim. GBank gij2143910[pri][S68216 - phosphatase-1 glycogen-binding (GL)-chain - rat]		phosphatase	264107, 264110, 264112, 265017, 263976
2020	94674476 (4039, 4040)	Novel Protein sim. GBank gij2078483 (U43200) - anilfreeze glycopeptide AFGP polypeptide precursor [Boreogadus salda]		UNCLASSIFIED	264259, 264508, 264591, 265018, 264682, 264288, 264688, 22278002
2021	86718818 (4041, 4042)	Novel Protein sim. GBank gij585084[sp][Q07803][EFGM_RAT - ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G)]			56994075, 264593, 33109954, 21906754, 21906768, 33657023, 33657109, 27486261, 87168518
2022	95295665 (4043, 4044)	Novel Protein sim. GBank gij4218005 (AC006135) - putative vicilin storage protein (globulin-like) [Arabidopsis thaliana]			264757, 264767, 60170615, 18108385
2023	87722976 (4045, 4046)	Novel Protein sim. GBank gij5410230[gb][A042992.1][AF07334 - (AF073344) ubiquitin specific protease 3 [Homo sapiens]	Contains protein domain (PF00442) - Ubiquitin carboxyl-terminal hydrolases family 2	ubiquitin	18108394, 22278999, 264259, 264905, 264906, 264908, 264595, 264762, 264769, 264634, 264636, 87168518, 60432113, 22279000, 264482, 264565
2024	87896443 (4047, 4048)				60433438, 265017, 264688, 264692, 264693, 264636
2025	87858863 (4049, 4050)	Novel Protein sim. GBank gij4755188[gb][A029055.1][AC00701 - (AC007018) unknown protein [Arabidopsis thaliana]	Contains protein domain (PF00637) - 7-fold repeat in Clathrin and VPS	UNCLASSIFIED	22278997, 264509, 264906, 264909, 55812038, 265017, 265021, 265022, 60170615, 264556
2026	94122114 (4051, 4052)	Novel Protein sim. GBank gij165569[emb][CAA69032] - (Y07752) perophorin-S [Volvox carteri]		UNCLASSIFIED	56994075, 60432049, 264508, 66712502, 264112, 60170831, 87168559, 264288, 264688, 264689, 21906768, 33657109, 18108370, 264638, 18108385, 60432113, 22279000, 22279002, 264564, 264566, 264567
2027	80249001 (4053, 4054)			UNCLASSIFIED	263978, 264634, 264486

2028	94699884 (4055, 4056)	Novel Protein sim. GBank gij1330345 (U58755) - coded for by C. elegans cDNA yk34b1.5; coded for by C. elegans cDNA yk13h10.5; coded for by C. elegans cDNA yk46e8.5; coded for by C. elegans cDNA yk46d5.5; coded for by C. elegans cDNA yk43c2.5; coded for by C. elegans cDNA yk46e8....	Contains protein domain (PF00632) - HECT-domain (ubiquitin-transferase).	ubiquitin	52644507, 52645156, 52646842, 56182575, 56994075, 35696286, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 66714117, 29331825, 29331826, 29331827, 29331828, 35696052, 264906, 264907, 264908, 29331830, 52644045, 56182435, 265006, 265009, 60432229, 33657402, 264595, 264757, 55812038, 21906754, 52646317, 52644296, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 18108354, 264288, 264389, 264766, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 52644150, 33657023, 65274620, 33657109, 52645129, 18108368, 27486261, 27486262, 27486264, 27486265, 35695763, 264629, 55811578, 35698423, 35695855, 264635, 264636, 52644332, 264558, 83373044, 56528486, 22278000, 22278002, 264563, 265009, 264595, 85658542, 264555, 264556, 264557, 264558, 264559, 83373044
2029	95362032 (4057, 4058)	Novel Protein sim. GBank gij3599940 (AF017368) - facio-genital dysplasia protein 2 [Mus musculus]	Contains protein domain (PF00621) - RhoGEF domain	UNCLASSIFIED	18108394, 56994075, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 264108, 66712502, 264828, 265009, 265018, 265019, 264681, 264682, 264684, 264685, 56181582, 264689, 21906769, 265022, 264692, 264693, 264628, 18108370, 18108374, 264634, 264636, 264556, 18108385, 87168518, 22279002, 264565
2030	91213734 (4059, 4060)	Novel Protein sim. GBank gij5630080 [gbjAAD5825.1]AC00489 - (AC004890) similar to HUB1; similar to BAA24380 (PID:g2789430) [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264591, 55811937, 18108365, 264557, 264558, 18108382, 18108384
2031	80245281 (4061, 4062)				
2032	91232607 (4063, 4064)	Novel Protein sim. GBank gij5689491 [dbjBAA83029.1] - (AB028000) KIAA1077 protein [Homo sapiens]	Contains protein domain (PF00884) - Sulfatase	hydrolase	65274572, 35696286, 29331824, 264908, 265009, 264593, 265018, 264288, 264686, 264769, 21906766, 21906767, 29148627, 264628, 35696423, 264634, 264556, 18108381, 60170394, 264559, 83373044, 18108385, 264482, 264484
2033	95000808 (4065, 4066)	Novel Protein sim. GBank gij2494828 [spQ64686]CAG7, RAT - ALPHA-N-ACETYL GALACTOSAMINIDE ALPHA-2,6-SIALYLTRANSFERASE (ST6GALNACIII) (STY)		synthase	56181562, 264628, 264632, 264555, 264556
2034	91232529 (4067, 4068)	Novel Protein sim. GBank gij4826984 [np_005147.1]PRD1 - UNKNOWN	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	22278996, 22278999, 264907, 29331830, 265008, 265018, 264681, 264682, 264684, 21906767, 21906768, 21906769, 33657109, 83373044, 56526486

2035	83553451 (4069, 4070)				264369, 264686, 265022, 56526486, 264567
2036	87115833 (4071, 4072)				29331827, 29331828, 264682, 264369, 29148627, 60432113
2037	943224833 (4073, 4074)	Novel Protein sim. GBank gi 2734081 (AF000195) - similar to oxysterol-binding proteins [Caenorhabditis elegans]		UNCLASSIFIED	65274572, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 29331824, 60432289, 29331826, 29331828, 35696052, 264907, 29331830, 66712502, 56182435, 265008, 265009, 60170831, 264594, 55812038, 33109954, 21906754, 87168559, 265017, 265018, 265019, 264762, 264369, 264288, 21906765, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 52644150, 33657023, 33657109, 33657182, 35695763, 35695855, 264632, 264634, 264636, 56182323, 83373044, 60432113, 22279000, 22279002, 264563
2038	95422384 (4075, 4076)	Novel Protein sim. GBank gi 3880625 embj CAB07858  (Z93785) predicted using GeneFinder; similar to RNA recognition motif (aka RRM, RBD, or RNP domain); cDNA EST EMBL.T01682 comes from this gene; cDNA EST EMBL.M75823 comes from this gene; cDNA EST EMBL.D27558 comes from this ge...	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	UNCLASSIFIED	22278995, 22278996, 56994075, 264259, 29331824, 35696052, 264905, 264906, 52644045, 265007, 265009, 87168559, 265017, 18108351, 264448, 264369, 264766, 264767, 264686, 18108358, 21908765, 21906769, 52644150, 33657023, 264692, 18108362, 33657109, 27486262, 18108370, 18108374, 18108379, 35696423, 65274791, 264632, 264636, 18108383, 83373044, 18108385, 87168518, 22279000, 22279002, 264563, 264564, 264566
2039	85514626 (4077, 4078)	Novel Protein sim. GBank gi 2224653 dbj BAA20813  (AB002354) KIAA0356 [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	22278997, 264259, 29331822, 264905, 264906, 264907, 264908, 264909, 264510, 265009, 264910, 264593, 264758, 265011, 265018, 264762, 264288, 264766, 264768, 264769, 21906766, 33657023, 264692, 264693, 33657109, 35696423, 264631, 264632, 264634, 264635, 264636, 264637, 264639, 87168518, 264488
2040	85308417 (4079, 4080)			UNCLASSIFIED	264592
2041	95071736 (4081, 4082)	Novel Protein sim. GBank gi 2500825 sp P70700 PRA2_MOUSE - DNA-DIRECTED RNA POLYMERASE I 135 KD POLYPEPTIDE (RNA POLYMERASE I SUBUNIT 2) (RPA135)	mapolymerase		264488, 22278998, 35696052, 264905, 264907, 264908, 264910, 265018, 264605, 265019, 18108351, 264766, 264769, 21906766, 265021, 265022, 264692, 33657109, 264628, 264629, 35696423, 35695855, 264637, 264638, 264563, 264564, 264565, 264567

2042	95307447 (4083, 4084)	Novel Protein sim. GBank gi 4406590 gb AAD20040  - (AF131766) Similar to Ena-VASP like protein [Homo sapiens]	Contains protein domain (PF00568) - WH1 domain	UNCLASSIFIED	60424179, 35696286, 264259, 29331826, 35696052, 29331828, 264508, 264509, 264907, 264909, 264510, 264511, 265009, 264910, 264591, 60433356, 264595, 265017, 265019, 264681, 264764, 264369, 264765, 264684, 264288, 264766, 264686, 52644229, 264769, 21906765, 35695917, 264535, 52644150, 264691, 264692, 18108385, 27486261, 27486262, 27486265, 18108374, 35696423, 65274791, 35695855, 264555, 264558, 60170394, 18108385, 264404, 22279000, 22279002, 264482, 264583, 264564, 264566
2043	94328076 (4085, 4086)	Novel Protein sim. GBank gi 5052554 gb AAD38607.1 AF14563 - (AF145632) BcDNA, GH06032 [Drosophila melanogaster]	Contains protein domain (PF00122) - E1-E2 ATPase	transport	264488, 52644507, 52646365, 56894075, 22278997, 22278999, 20281171, 264259, 29331822, 29331824, 66714117, 29331826, 29331828, 33656970, 29146498, 264509, 264908, 52644045, 56182435, 265006, 33657402, 21906754, 52644296, 87168559, 265017, 265018, 265019, 264681, 264288, 264766, 264685, 264686, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 60170615, 264691, 33657023, 264693, 65274620, 33657109, 33657182, 27486261, 27486262, 33657349, 35695763, 18108374, 55811576, 35695855, 18108380, 18108381, 60170394, 56182323, 264558, 83373044, 18108385, 56526486, 87168518, 60432113, 22279000, 264567
2044	87106927 (4087, 4088)	Novel Protein sim. GBank gi 2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		struct	264093, 29331827, 264905, 66712502, 264592, 264689, 21906765, 21906769, 265020, 264692, 264482, 264566
2045	79635532 (4089, 4090)				264692
2046	87320849 (4091, 4092)	Novel Protein sim. GBank gi 4406698 gb AAD20062  - (AF131852) Unknown [Homo sapiens]			264259, 264908, 264683, 22279002
2047	84578801 (4093, 4094)	Novel Protein sim. GBank gi 4101720 (AF006466) - lymphocyte specific formin related protein [Mus musculus]			22278999, 29147620, 29331824, 29146498, 264508, 265007, 265008, 265019, 264605, 264681, 29148627, 29148629, 265021, 33657023, 18108365, 33657109, 33657182, 18108377, 264556, 264638, 264559, 18108388
2048	84606378 (4095, 4096)			UNCLASSIFIED	264909
2049	88094690 (4097, 4098)	Novel Protein sim. GBank gi 4589656 gb BAA76850.1  - (AB023223) KIAA1006 protein [Homo sapiens]		UNCLASSIFIED	264488, 264259, 29331824, 29331826, 35696052, 264906, 264907, 264908, 264909, 264910, 264603, 264763, 21906767, 21906768, 264629, 264634, 264637, 22279002, 264564, 264565, 264566, 264567



2050	79633835 (4089, 4100)				264693	264488, 264259, 264509, 264906, 264907, 264769, 18108374, 35696423, 264563, 264566, 264486
2051	87780168 (4101, 4102)				UNCLASSIFIED UNCLASSIFIED	
2052	88096393 (4103, 4104)	Novel Protein sim. GBank gi 4529889 gb AAD21812.1  - (AF134726) G9A [Homo sapiens]	Contains protein domain (PF00856) - SET domain	kinase	UNCLASSIFIED	264488, 263994, 35696052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264758, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33657023, 33657109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264564, 264566, 264488, 264567, 264488, 263994, 35696052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264758, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 33657023, 33657109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264564, 264566, 264488, 264567
2053	87763078 (4105, 4106)	Novel Protein sim. GBank gi 2995449 emb CAA75113  - (Y14848) midline 1 protein [Mus musculus]			UNCLASSIFIED	22278996, 22278997, 264259, 29331822, 264102, 264508, 35695917, 263972, 264482
2054	95358937 (4107, 4108)	Novel Protein sim. GBank gi 3876326 emb CAB02090  - (Z79754) similar to C2 domain [Caenorhabditis elegans]	Contains protein domain (PF00168) - C2 domain			60424179, 264094, 264259, 29331825, 60424269, 264908, 60432229, 60433356, 87168559, 265019, 264760, 264288, 264686, 21906769, 33657023, 264693, 55810764, 55811576, 264635, 56182323, 60432113
2055	88259449 (4109, 4110)	Novel Protein sim. GBank gi 5353746 gb AAD42226.1 AF15913 - (AF159133) SIR2-like protein [Oryza saliva subsp. indica]			UNCLASSIFIED	264488, 29331826, 60432289, 29331828, 60433356, 265019, 264683, 264684, 265021, 33657109, 18108374, 264637, 18108385, 87168518, 60432113, 22279000, 264564

2056	88177398 (4111, 4112)	Novel Protein sim. GBank gij4826960[ref]NP_005042.1[pOARS - glutamine-tRNA synthetase	Contains protein domain (PF00749) - tRNA synthetases class I (E and Q)	synthase	264488, 52645156, 56182575, 22278994, 35696286, 56994075, 22278998, 22278998, 22278999, 60432049, 264259, 29331824, 60432289, 29331827, 29331828, 33656970, 264104, 264908, 264908, 265006, 265008, 60170831, 264591, 60432229, 60433438, 18108348, 21906754, 33657084, 52644296, 87168474, 265010, 87168559, 265017, 265018, 264760, 18108351, 264681, 264682, 264448, 264683, 264369, 264288, 264685, 264687, 264688, 264689, 21908765, 21908766, 21908767, 21908769, 55811957, 35695917, 265022, 33657023, 18108362, 33657109, 18108368, 33657182, 27486261, 27486264, 27486265, 33657349, 264628, 18108370, 264629, 18108374, 18108377, 18108379, 35696423, 55811576, 20281152, 264636, 264952, 18108385, 18108388, 87168518, 264482, 264565, 264566, 264567, 52646842, 52646365, 56182575, 35696286, 22278996, 22278997, 22278998, 264093, 52645080, 35696052, 29331828, 33656970, 265009, 52646317, 55811386, 52644296, 52644229, 21906769, 35695917, 265021, 60170615, 52644150, 33657109, 33657182, 27486261, 27486262, 35695763, 35696423, 35695855, 52644332
2057	87877905 (4113, 4114)	Novel Protein sim. GBank gij728850[sp]P08640[AMYH_YEAST - GLUCOAMYLASE S1/S2 PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)		UNCLASSIFIED	265007, 265008, 264591
2058	86276896 (4115, 4116)	Novel Protein sim. GBank		UNCLASSIFIED	29331825, 264682, 264686, 264691, 264693, 22278002
2059	79866684 (4117, 4118)	gij119714[sp]P13983[EXTN_TOBAC - EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)]		UNCLASSIFIED	
2060	83050800 (4119, 4120)	Novel Protein sim. GBank gij2811122 (U87318) - NaDC-2 [Xenopus laevis]		UNCLASSIFIED	56182575, 29331824, 29331826, 264910, 55811957, 18108370, 55811576

2061	95362204 (4121, 4122)	Novel Protein sim. GBank gi 2496947 sp Q09298 YOO9_CAEEL - HYPOTHETICAL 141.2 KD PROTEIN EEED8.9 IN CHROMOSOME II	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	22278997, 22278998, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 29331828, 264906, 66712502, 29331830, 264908, 264909, 264112, 264511, 265007, 265009, 264910, 264591, 33657402, 21906754, 85658542, 265017, 265019, 264448, 264683, 264288, 264684, 264369, 264686, 264687, 18108358, 264689, 21906765, 21906767, 21906768, 21906769, 265020, 265022, 264691, 33657023, 33657109, 20281149, 18108379, 35695855, 264634, 264556, 264557, 264558, 18108382, 264559, 83373044, 18108384, 56528486, 60432113
2062	87028440 (4123, 4124)	Novel Protein sim. GBank gi 4502081 ref NP_001139.1 pANK2 - ankyrin 2, neuronal	Contains protein domain (PF00023) - struct Ank repeat	264905, 264628, 264907, 264629, 264908, 264909, 18108374, 263978, 35695855, 264512, 264635, 60431850, 264636, 264760, 264563, 18108351, 264762, 264565, 264764, 264487, 264766
2063	87601272 (4125, 4126)	Novel Protein sim. GBank gi 4589562 dbj BAA78803.1  - (AB023178) KIAA0959 protein [Homo sapiens]	Contains protein domain (PF00617) - oncogene RasGEF domain	22278994, 22278999, 264259, 29331827, 264906, 264909, 52644045, 264686, 21906767, 55811957, 264692, 18108365, 263972, 55811576, 18108384, 22279002, 264482, 264563, 264564, 264484
2064	95317253 (4127, 4128)	Novel Protein sim. GBank gi 1754515 dbj BAA13413.1  - (D87515) aminopeptidase-B [Rattus norvegicus]	hydrolase	264488, 52646365, 56994075, 35696286, 22278997, 22278998, 264259, 29331828, 60432289, 29331827, 29331828, 35696052, 264509, 265007, 265008, 60432229, 60433438, 21906754, 265010, 265011, 87168559, 265017, 265018, 264761, 18108351, 264682, 264369, 264288, 52644229, 21906765, 21906767, 21906768, 35695917, 33657109, 18108368, 18108374, 35698423, 35695855, 52644332, 264559, 60432113, 22279000, 22279002, 264566, 264486
2065	95092238 (4129, 4130)	Novel Protein sim. GBank gi 2507144 sp Q04205 TENS_CHICK - TENSIN	kinase	264569, 18108394, 56182181, 60432289, 29331826, 264905, 264906, 264908, 60431735, 60433358, 55811388, 85658542, 265018, 55811150, 264681, 264766, 264692, 60431528, 263974, 55810764, 35695855, 264631, 264634, 264635, 60431850, 264557, 83373044, 18108388, 22279000, 22279002, 56182575, 264259, 264906, 264764, 264288, 56182323, 264567
2066	85783402 (4131, 4132)	Novel Protein sim. GBank gi 160171 (M58295) - circumsporozoite protein [Plasmodium yoelii]	Contains protein domain (PF00096) - UNCLASSIFIED Zinc finger, C2H2 type	

2067	95303892 (4133, 4134)				35696286, 22278997, 22278998, 60432049, 264259, 60432289, 60433438, 264682, 264448, 264369, 264288, 18108355, 21906765, 21906768, 265022, 33657109, 35696423, 35695855, 264558, 264404, 264563, 264486
2068	94344754 (4135, 4136)			UNCLASSIFIED	264687
2069	94319177 (4137, 4138)	Novel Protein sim. GBank gi 3152662 (AF064604) - KE03 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	transcriptfactor	60424170, 56182575, 22278995, 22278996, 56994075, 264259, 29331822, 29331824, 29331825, 35696052, 29331828, 33658970, 264509, 264905, 56182435, 265009, 60433356, 87168559, 265017, 265018, 264604, 265019, 264448, 264764, 264766, 21906765, 21906767, 21906768, 21906769, 265020, 265021, 33657023, 33657109, 263976, 264555, 264557, 56182323, 83373044, 87168518, 60432113, 22279000, 22279002
2070	85791380 (4139, 4140)	Novel Protein sim. GBank gi 5712131 gb AAD47379.1 AF12049 - (AF120499) DEM1 protein [Homo sapiens]		UNCLASSIFIED	35695917, 264905, 284628, 264908, 264638
2071	86946116 (4141, 4142)	Novel Protein sim. GBank gi 3551531 dbj BAA33016  - (AB017437) avens [Gallus gallus]	Contains protein domain (PF00568) - WH1 domain		18108398, 265006, 265007, 265008, 265009, 264594, 265010, 265011, 18108351, 18108354, 18108364, 18108365, 18108368, 264634, 18108381, 18108385, 18108388, 18108391
2072	91718428 (4143, 4144)		Contains protein domain (PF00184) - Neurohypophysis hormones, C-terminal Domain		22278995, 35696286, 22278997, 22278998, 22278999, 264490, 60432049, 264259, 29331822, 29331824, 29331828, 35696052, 265008, 33657402, 21906754, 265011, 265019, 18108351, 264682, 264369, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 264690, 264692, 35696423, 264555, 264556, 264558, 22279000
2073	27925684 (4145, 4146)	Novel Protein sim. GBank gi 1504026 dbj BAA13212  - (D86976) similar to C.elegans protein (Z37093) [Homo sapiens]		UNCLASSIFIED	264556
2074	94324767 (4147, 4148)	Novel Protein sim. GBank gi 4240317 dbj BAA74937.1  - (AB020721) KIAA0914 protein [Homo sapiens]			28331822, 264909, 264511, 265008, 264594, 264595, 265010, 265011, 265017, 265018, 265019, 264448, 264683, 265020, 265021, 18108370, 264632, 83373044, 264567

2075	94314886 (4149, 4150)	Novel Protein sim. GBank gij5138930[gbjAAD40382.1] - (AF093680) transcription factor IIB [Homo sapiens]			18108394, 22278994, 22278996, 35656286, 22278998, 22278999, 264259, 29331822, 29331825, 29331827, 35696052, 29331828, 264905, 264907, 264908, 264510, 265007, 264910, 265009, 33657084, 264760, 264448, 264288, 264766, 264767, 264689, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 33657023, 27486262, 27486265, 35696423, 35695855, 264631, 264634, 264639, 87168518, 22279002, 264563, 264486, 18108391
2076	87594118 (4151, 4152)				264259, 29331826, 264508, 264908, 264510, 265007, 265011, 264288, 264637, 18108385
2077	11389877 (4153, 4154)			UNCLASSIFIED	264592
2078	87539364 (4155, 4156)	Novel Protein sim. GBank gij4220590[dbjBAA74579] - (D87908) nuclear protein np95 [Mus musculus]	Contains protein domain (PF00628) - PHD-finger	ubiquitin	29331825, 265017, 265018, 264288, 265020, 265021, 264634, 56526486
2079	88095916 (4157, 4158)	Novel Protein sim. GBank gij4240255[dbjBAA74906.1] - (AB020690) KIAA0883 protein [Homo sapiens]	Contains protein domain (PF00098) - Zinc finger, CCHC class	UNCLASSIFIED	264768, 22278997, 265021, 264690, 264259, 264692, 29331822, 264693, 29331824, 29331828, 264508, 264509, 264908, 264907, 264628, 20281069, 264909, 265007, 265009, 264632, 264636, 264591, 264592, 264639, 264758, 264759, 33109954, 264604, 265018, 265019, 22279002, 264563, 264564, 264448, 264684, 264567, 264885
2080	94136689 (4159, 4160)	Novel Protein sim. GBank gij2408021[embjCAB16219.1] - (Z99162) putative vacuolar protein [Schizosaccharomyces pombe]	Contains protein domain (PF01363) - FYVE zinc finger	UNCLASSIFIED	56182575, 264092, 29331824, 29331826, 29331830, 265017, 265018, 265020, 83373044
2081	94847186 (4161, 4162)	Novel Protein sim. GBank gij5524734[gbjAAD44360.1]AF16635 - (AF166350) ST7 protein [Homo sapiens]	Contains protein domain (PF00431) - eph CUB domain	eph	22278996, 22278997, 22278999, 60432049, 29331822, 29331824, 33657402, 85658542, 265011, 265018, 265019, 21906787, 21906788, 264693, 18108385, 22279000, 22279002
2082	87628629 (4163, 4164)	Novel Protein sim. GBank gij3880558[embjCAA94234] - (Z70271) predicted using Genefinder; similar to collagen; cDNA EST yk308e7.3 comes from this gene; cDNA EST yk308e7.5 comes from this gene; cDNA EST yk385a8.3 comes from this gene; cDNA EST yk385a8.5 comes from this gene [Caeno...]		collagen	264907, 265019
2083	94141000 (4165, 4166)	Novel Protein sim. GBank gij2352427 (AF004161) - peroxisomal Ca-dependent solute carrier [Cryptosporidium parvum]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	55182575, 22278996, 22278997, 264259, 29331822, 29331825, 264509, 264112, 265009, 264593, 21906754, 265018, 265019, 264448, 264288, 264685, 21906769, 265020, 265022, 264691, 18108370, 65274791, 264631, 264555, 264563
2084	95199298 (4167, 4168)	Novel Protein sim. GBank gij728836[spP39193]ALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII		UNCLASSIFIED	265018, 264763, 264683, 264691

2085	94989476 (4169, 4170)	Novel Protein sim. GBank gi 1655699 emb CAA69032  - (Y07752) perophorin-S [Voivox carten]		UNCLASSIFIED	56182575, 60432289, 264908, 56182435, 87168474, 264763, 264369, 264688, 264693, 18108370, 56182323
2086	91234404 (4171, 4172)	Novel Protein sim. GBank gi 3875032 emb CAA88936  - (Z49125) similarity to Trichostrogylus colubriformis 11 kd secretory protein (Swiss Prot accession number P21837); cDNA EST EMBL:D33348 comes from this gene; cDNA EST EMBL:D37644 comes from this gene; cDNA EST EMBL:D36149 come....		UNCLASSIFIED	35696286, 264259, 35696052, 264906, 264907, 264908, 264909, 264910, 264759, 264604, 264762, 264768, 264769, 35695917, 263978, 35698423, 35695855, 264632, 264634, 264637, 264638, 264639, 56182323, 18108385, 264482, 264486
2087	21436337 (4173, 4174)			UNCLASSIFIED	264489
2088	94111527 (4175, 4176)	Novel Protein sim. GBank gi 3880930 emb CAA16334.1  - (AL021481) similar to Phosphoglucosylase and phosphomannomutase phosphoserine; cDNA EST EMBL:D36168 comes from this gene; cDNA EST EMBL:D70697 comes from this gene; cDNA EST yk373h9.5 comes from this gene; cDNA EST EMBL:T0080....			264488, 22278994, 35696286, 22278996, 29331827, 35696052, 33657402, 21908754, 33109954, 87168474, 265017, 265018, 285019, 264448, 264683, 264369, 264685, 264687, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264692, 33657023, 33657109, 33657182, 27486261, 27486262, 33657348, 27486265, 35698423, 35695855, 83373044, 87168518, 22279000, 264567
2089	95422801 (4177, 4178)	Novel Protein sim. GBank gi 4758118 ref NP_004623.1 pDAP3 - Death associated protein 3		cadherin	18108392, 264488, 52644507, 18108394, 18108397, 52648842, 18108398, 56182575, 22278994, 22278995, 35696286, 22278996, 56994075, 22278997, 22278998, 22278999, 264091, 264092, 264093, 264094, 60432049, 264259, 29331822, 20281099, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 33656970, 29146498, 29146499, 264102, 264106, 264107, 264109, 264508, 264905, 264509, 264906, 264907, 264908, 66712502, 264828, 52644045, 264809, 56182435, 264110, 264112, 264510, 264511, 265006, 264512, 265007, 265008, 264910, 265009, 60170831, 264592, 264593, 60433356, 33657402, 80433438, 264595, 55812038, 264758, 21906754, 33657084, 55811386, 52644296, 265010, 265011, 87168559, 265017, 265018, 265019, 264760, 264761, 55811150, 264762, 18108351, 264682, 264448, 264763, 264764, 264683, 264369, 18108354, 264288, 264685, 264766, 264686, 264687, 264768, 52644229, 264688, 18108358, 56181562, 264769, 18108359, 264689, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 55811957, 29148629, 29148784, 35695917, 265020, 265021, 265022, 60170615, 264690

2090	88222470 (4179, 4180)				22278995, 22278998, 22278999, 264259, 22278999, 22278999, 264259, 60433438, 33109954, 87168474, 87168559, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906769, 35695917, 265022, 60170615, 33657023, 35696423, 35695855, 264952, 18108387, 22279000, 263994, 264805, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906766, 21906769, 265022, 264628, 264563, 264567
2091	95309161 (4181, 4182)	Novel Protein sim. GBank gij4580997/gb/AAD24571.1JAF12108 - (AF121081) cAMP inducible 2 protein [Mus musculus]		UNCLASSIFIED	22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2092	88223605 (4183, 4184)			homeobox	22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2093	87406073 (4185, 4186)	Novel Protein sim. GBank gij2352427 (AF004161) - peroxisomal Ca-dependent solute carrier [Oryctolagus cuniculus]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	264905, 264906, 264907, 264908, 264510, 265006, 265007, 265009, 264910, 264596, 21906754, 87168474, 265011, 264603, 265018, 265019, 264760, 264766, 264768, 264769, 21906767, 21906768, 21906769, 265021, 264690, 33657023, 264693, 264628, 264634, 264636, 264637, 264557, 56182323, 264564
2094	91230928 (4187, 4188)	Novel Protein sim. GBank gij4929551/gb/AAD34036.1JAF15179 - (AF151799) CGI-40 protein [Homo sapiens]		MHC	35696286, 265017, 265018, 265019, 18108388
2095	95351526 (4189, 4190)	Novel Protein sim. GBank gij1363238/jprj/A57284 - spermatid perinuclear RNA-binding protein Spnr - mouse	Contains protein domain (PF00035) - Double-stranded RNA binding motif	dna_ma_bind	35696286, 52644045, 265006, 265007, 265008, 87168559, 18108351, 21906769, 29148784, 265020, 33657023, 27486262, 18108374, 18108388
2096	94119760 (4191, 4192)	Novel Protein sim. GBank gij3834423 (AF070689) - cytoplasmic dynein intermediate chain isoform DIC1a [Drosophila melanogaster]	Contains protein domain (PF00400) - WD domain, G-beta repeat	ATPase_associated	264488, 264489, 65274572, 56182575, 22278996, 22278997, 22278999, 264259, 60432289, 29331826, 35696052, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 52644045, 264909, 264510, 264511, 264512, 265008, 264910, 265009, 264592, 60433356, 60433438, 264758, 264596, 55812038, 21906754, 264601, 264602, 264605, 264762, 264681, 18108351, 264764, 264683, 264288, 264687, 264768, 264769, 264689, 21906765, 21906766, 21906767, 35695917, 265020, 265022, 52644150, 264691, 264692, 33657023, 264693, 27486281, 35695763, 264628, 264629, 35696423, 35695855, 264631, 264632, 264634, 264635, 264555, 264637, 263981, 264638, 264639, 264563, 264483, 264565, 264566, 264486, 264567

2097	95322772 (4193, 4194)	Novel Protein sim. GBank gi 5174501 ref NP_006051.1 pLYF1 - zinc finger protein, subfamily 1A, 1 (Ikaros)	(Contains protein domain (PF000096) - Zinc finger, C2H2 type	transcriptfactor	65274572, 264511, 265010, 264600, 265017, 264448, 264288, 265021, 60170815, 264692, 33657109, 18108370, 264636, 264483 56994075, 264259, 264288, 265020, 264563
2098	87780340 (4195, 4196)	Novel Protein sim. GBank gi 4758208 ref NP_004081.1 pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	phosphatase	
2099	95412927 (4197, 4198)	Novel Protein sim. GBank gi 2695659 (AF026954) - pyruvate dehydrogenase phosphatase regulatory subunit precursor, PDP1 [Bos taurus]		phosphatase	65274572, 264905, 65274444, 264691, 264636, 264555
2100	95332656 (4199, 4200)	Novel Protein sim. GBank gi 3881189 emb CAB16514  - (Z99281) similar to ADP-ribosylation factor; cDNA EST EMBL: C08179 comes from this gene; cDNA EST EMBL: C08337 comes from this gene; cDNA EST EMBL: C09829 comes from this gene; cDNA EST yk291b4.5 comes from this gene; cDNA EST yk4...	Contains protein domain (PF00025) - ADP-ribosylation factor family	nuc1_recp	56182575, 22278995, 22278996, 22278997, 22278998, 60432049, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 29146498, 264909, 265008, 265009, 264910, 284591, 60432229, 60433356, 33657402, 264758, 21908754, 85658542, 87168474, 285017, 265018, 265019, 264681, 18108351, 264762, 264448, 264369, 264288, 18108355, 264686, 21908765, 21908767, 21908768, 21908769, 265020, 265021, 33657023, 18108374, 35696423, 264558, 83373044, 87168518, 60432113, 22279000, 22279002 264091, 29331824, 264105, 265007, 265010, 18108380
2101	87762604 (4201, 4202)	Novel Protein sim. GBank gi 4589468 dbj BAA76761.1  - (AB012808) mBOCT [Mus musculus]		UNCLASSIFIED	
2102	87770461 (4203, 4204)	Novel Protein sim. GBank gi 3874149 emb CAA97423.1  - (Z73103) predicted using GeneFinder [Caenorhabditis elegans]		UNCLASSIFIED	264488, 264489, 35696286, 264259, 35696052, 264508, 264905, 264907, 264908, 264909, 264511, 264512, 264591, 264593, 60433356, 264758, 264601, 264605, 264760, 18108351, 264448, 264764, 264288, 264767, 264768, 21908769, 35695917, 18108374, 264634, 264555, 264559, 264563, 264482, 264486
2103	95413576 (4205, 4206)	Novel Protein sim. GBank gi 4240159 dbj BAA74858.1  - (AB020842) KIAA0835 protein [Homo sapiens]	Contains protein domain (PF01530) - Zinc finger, C2HC type	transcriptfactor	65274572, 56994075, 22278999, 264259, 29331824, 29331825, 35696052, 29331828, 66712502, 265009, 60170831, 264595, 33109954, 85658542, 87168559, 265017, 265019, 264448, 21908765, 21908768, 265022, 33657023, 27486262, 33657349, 35695763, 60431528, 18108374, 55811576, 56182323, 18108387, 87168518, 60432113, 264564
2104	85776161 (4207, 4208)			UNCLASSIFIED	264592, 264604, 22279000



2105	94848080 (4209, 4210)	Novel Protein sim. GBank gij1707032 (U80445) - coded for by C. elegans cDNA yk13g5.3; coded for by C. elegans cDNA yk21g6.3; coded for by C. elegans cDNA CEMSE18F; coded for by C. elegans cDNA yk126b1.3; coded for by C. elegans cDNA yk65h8.3; coded for by C. elegans cDNA yk65h8.....		UNCLASSIFIED	264488, 56182575, 22278994, 56894075, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 29331828, 264508, 264905, 264509, 264907, 29331830, 52844045, 264510, 264511, 265007, 264512, 265009, 60170831, 60432229, 33657402, 60433358, 264595, 60433438, 264758, 33657084, 87168474, 265010, 87168559, 265017, 265018, 265019, 264762, 18108351, 264684, 18108354, 264288, 264686, 52644229, 18108359, 21808765, 21906766, 21906767, 21908788, 21908789, 35695917, 265020, 265022, 60170615, 52644150, 264692, 33657023, 264693, 33657109, 60431528, 18108374, 65274791, 35695855, 264635, 60170394, 264639, 264558, 18108385, 18108387, 56526486, 87168518, 60432113, 264564, 264566, 264567, 265006, 265019
2106	83365475 (4211, 4212)				264906, 264639
2107	78822662 (4213, 4214)	Novel Protein sim. GBank gij3881524[embjCAA93883] - (Z70038) ZK1067.4 [Caenorhabditis elegans]			
2108	94233976 (4215, 4216)	Novel Protein sim. GBank gij3176689 (AC003671) - Contains similarity to ubiquitin carboxyl-terminal hydrolase 14 gbZ35927 from S. cerevisiae. [Arabidopsis thaliana]		UNCLASSIFIED	264905, 264906, 264907, 264908, 264909, 264758, 265011, 264600, 264601, 264784, 264766, 264767, 264768, 264769, 264693, 264629, 35695855, 264632, 264634, 264635, 264638, 264639, 83373044, 264486
2109	80478719 (4217, 4218)			UNCLASSIFIED	18108348, 264769, 18108370, 18108374, 264555, 264556, 264557, 264558
2110	87729075 (4219, 4220)	Novel Protein sim. GBank gij481043[pirjS37671] - bat2 protein - human		UNCLASSIFIED	264564
2111	87818419 (4221, 4222)			UNCLASSIFIED	264766, 35695917, 264630, 264567, 264486
2112	87293783 (4223, 4224)	Novel Protein sim. GBank gij2143639[pirjI56542] - calmodulin-binding protein - rat	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	struct	264508, 264906, 264591, 264662, 22279002
2113	78941388 (4225, 4226)	Novel Protein sim. GBank gij4426629[gbjAAD20459] - (AF100960) protocadherin [Rattus norvegicus]	Contains protein domain (PF00028) - Cadherin domain	cadherin	265006
2114	87889342 (4227, 4228)	Novel Protein sim. GBank gij3327184[dbjBAA31660] - (AB014585) KIAA0685 protein [Homo sapiens]			66714117, 29331826, 29331827, 60433438, 55812038, 265017, 265019, 264689, 21906769, 55811957, 265020, 265021, 33657109, 60170394, 264558
2115	90893785 (4229, 4230)	Novel Protein sim. GBank gij4757890[refjNP_004328.1]pC8OR - chromosome 8 open reading frame 1		UNCLASSIFIED	65274572, 264689, 264691, 264692, 60432113

2116	88259387 (4231, 4232)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		struc	35696286, 22278999, 56182181, 29331824, 29331825, 29331827, 35696052, 264907, 56182435, 265008, 264591, 55812038, 55811386, 87168559, 264288, 264369, 21906769, 29148629, 33657023, 35695763, 55811576, 35696423, 18108385
2117	87788904 (4233, 4234)	Novel Protein sim. GBank gij2330021 (AF019250) - kinesin-related protein; KRP; Costal2 [Drosophila melanogaster]		struc	29331824, 264511, 265009, 33109954, 265017, 265018, 264288, 264689, 265020, 264692, 58526486, 264482
2118	87078894 (4235, 4236)	Novel Protein sim. GBank gij1078307 [pji]B56573 - nuclear pore complex glycoprotein p62 - African clawed frog		glycoprotein	264259, 264905, 264907, 264908, 264510, 264511, 265009, 264910, 265010, 264602, 264288, 264768, 264693, 263967, 263972, 264638, 264559
2119	86999317 (4237, 4238)	Novel Protein sim. GBank gij4321407 [gb]AAD15748] - (AF047690) ATP-binding cassette protein M-ABC1 [Homo sapiens]		UNCLASSIFIED	264693, 18108385
2120	87789395 (4239, 4240)	Novel Protein sim. GBank gij4885527 [ref]NP_005480.1 [pNSP3 - novel SH2-containing protein 3	Contains protein domain (PF00017) - Src homology domain 2	eph	264091, 264259, 29331826, 29331828, 265017, 264604, 264288, 264685, 265020, 264691, 18108370, 55810764, 264555, 264636, 60432113
2121	80021375 (4241, 4242)	Novel Protein sim. GBank gij4757728 [ref]NP_004886.1 [pAGTA - angiotensin/vasopressin receptor All/AVP-like		UNCLASSIFIED	264601, 264766, 263978
2122	91230931 (4243, 4244)	Novel Protein sim. GBank gij4929551 [gb]AAD34036.1 [AF15179 - (AF151799) CGI-40 protein [Homo sapiens]			18108394, 56182575, 22278997, 29331822, 29331824, 29331825, 29331826, 29331828, 264907, 56182435, 265007, 264910, 265010, 265018, 264686, 265020, 55811576, 264555, 264637, 18108382, 83373044, 18108383, 18108384, 56528486, 264565, 264567
2123	86787998 (4245, 4246)	Novel Protein sim. GBank gij2224551 [dbj]BAA20764] - (AB002303) KIAA0305 [Homo sapiens]	Contains protein domain (PF01363) - FYVE zinc finger	struc	18108396, 264757, 265011, 18108351, 264691, 264634, 18108385
2124	83005951 (4247, 4248)	Novel Protein sim. GBank gij5689455 [dbj]BAA83011.1] - (AB028882) KIAA1059 protein [Homo sapiens]	Contains protein domain (PF00801) - PKD domain	transport	29331822, 264906, 264907, 264591, 264639, 264563
2125	95354041 (4249, 4250)	Novel Protein sim. GBank gij728831 [sp]P39188 [ALU1_HUMAN - III] ALU SUBFAMILY J WARNING ENTRY III]		UNCLASSIFIED	264259, 264509, 264907, 264511, 85658542, 264763, 21906785, 35695917, 264636, 264488
2126	95084231 (4251, 4252)	Novel Protein sim. GBank gij4539264 [emb]CAB39853.1] - (AL049495) conserved hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	264488, 264489, 29331827, 35696052, 264905, 264509, 264908, 264909, 264510, 265009, 264591, 264592, 264593, 33657402, 264594, 264595, 264596, 264758, 264601, 264603, 265018, 264604, 264605, 264760, 264681, 264762, 264683, 264764, 264684, 264288, 264685, 264689, 60170615, 33657023, 33657109, 55810764, 264635, 264636, 264637, 264638, 264639, 83373044, 264564, 264566

2127	81118652 (4253, 4254)	Novel Protein sim. GBank gi 4668435 gb AAD31315.1 AF14323 - (AF143236) apoptosis related protein APR-2 [Homo sapiens]			35696286, 29331826, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265006, 264511, 264512, 265007, 265009, 264910, 264758, 265011, 264600, 264601, 264604, 264762, 264763, 264766, 264687, 264768, 264769, 264689, 35695917, 264690, 264691, 264692, 264693, 264629, 18108374, 35695855, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 18108385, 264563, 264564, 264567, 56182575, 35696286, 56182181, 29331824, 60432289, 35696052, 264905, 264907, 66712502, 264908, 264909, 264510, 264512, 265009, 264910, 264591, 55812038, 265018, 264764, 264288, 264389, 264687, 264768, 55811957, 264692, 18108368, 264628, 264632, 264634, 264635, 264637, 56182323, 264639, 18108384, 18108388, 264563, 264567
2128	87414262 (4255, 4256)				66714117, 264828, 264595, 55812038, 55811150, 55811957, 264693, 18108374, 263978, 85274791, 18108381, 83373044, 22279000
2129	95102089 (4257, 4258)		UNCLASSIFIED		263981
2130	95417144 (4259, 4260)	Novel Protein sim. GBank gi 2649255 (AE001012) - conserved hypothetical protein [Archaeoglobus fulgidus]	UNCLASSIFIED		35696052, 264909, 264768, 35695917
2131	85723065 (4261, 4262)	Novel Protein sim. GBank gi 1086886 (U41276) - Similar to potassium channel protein. [Caenorhabditis elegans]	potassium_channel	Contains protein domain (PF00805) - Pentapeptide repeats (8 copies)	264488, 22278999, 264259, 29331827, 29331828, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265006, 264511, 265007, 284910, 264591, 264592, 264595, 264758, 21906754,
2132	95361096 (4263, 4264)	Novel Protein sim. GBank gi 5689373 dbj BAA82973.1  - (AB028944) KIAA1021 protein [Homo sapiens]	ATPase_associated	Contains protein domain (PF00122) - E1-E2 ATPase	33109954, 87168474, 265011, 264600, 264601, 264605, 265019, 264760, 18108351, 264681, 264762, 264764, 264288, 264684, 264766, 264686, 264687, 264768, 264769, 264688, 21908769, 264690, 52644150, 264691, 264693, 18108370, 264628, 264629, 18108372, 18108374, 35696423, 35695855, 264631, 264634, 264635, 264636, 264555, 264637, 18108380, 264639, 264558, 56182323, 56528486, 264564, 264565, 264566, 264567

2133	95351539 (4265, 4266)	Novel Protein sim. GBank gij4220489 (AC006069) - hypothetical protein [Arabidopsis thaliana]	UNCLASSIFIED	60424179, 52646365, 52646842, 56994075, 35696286, 22278997, 22278998, 60432049, 56182181, 66714117, 60424269, 29331826, 29331828, 35696052, 264905, 264906, 264907, 66712502, 29331830, 56182435, 265006, 264512, 265008, 60431735, 60433356, 33657402, 55812038, 33109954, 21906754, 55811386, 265010, 264603, 265017, 265018, 265019, 55811150, 18108351, 264682, 264369, 264288, 52644229, 56181562, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 33657023, 33657109, 60431528, 18108374, 35696423, 65274791, 35695855, 264634, 60431850, 18108380, 56182323, 83373044, 18108385, 18108387, 60432113, 222789002, 264563, 264568
2134	95412697 (4267, 4268)	Novel Protein sim. GBank gij3875351[emb]CAB08415[- (Z96047) DY3.6 [Caenorhabditis elegans]		55181686, 35696286, 21906754, 55811386, 265011, 265017, 18108351, 264765, 264766, 264688, 21906768, 35695917, 265020, 33657023, 264628, 35695855, 264632, 264555, 264556, 264557, 264558, 18108382, 222789002
2135	86078813 (4269, 4270)	Novel Protein sim. GBank gij5689559[dbj]BAA83063.1[- (AB029034) KIAA1111 protein [Homo sapiens]	UNCLASSIFIED	22278998, 29331828, 35696052, 264906, 264908, 264910, 265009, 264591, 264758, 52646317, 265011, 87168559, 264601, 18108351, 264448, 264683, 264684, 264689, 18108359, 264691, 33657023, 264692, 35695763, 264629, 35695855, 264631, 264635, 264636, 264637, 56182323, 264639, 222789002, 264564
2136	84346479 (4271, 4272)	Novel Protein sim. GBank gij2662167[dbj]BAA23715[- (AB007903) KIAA0443 [Homo sapiens]	UNCLASSIFIED	264639
2137	87637716 (4273, 4274)	Novel Protein sim. GBank gij4884110[emb]CAB43262.1[- (AL050090) hypothetical protein [Homo sapiens]	UNCLASSIFIED	264569, 264909, 33109954, 264763, 21906768, 60170394, 18108385, 264563
2138	87395446 (4275, 4276)	Novel Protein sim. GBank gij5174779[gb]AAD0696.1[- (U87804) 50 kDa protein [Caulobacter crescentus]	ATPase_associated	264259, 29331828, 35696052, 264909, 265006, 265017, 265018, 18108351, 264288, 21906768, 33657023, 33657109, 264628, 18108374, 35695855, 264634, 264555, 264556, 264557, 264558, 264559
2139	94843882 (4277, 4278)	Novel Protein sim. GBank gij3850821[emb]CAA77135[- (Y18350) U2 snRNP auxiliary factor, large subunit [Nicotiana plumbaginifolia]		264905, 264910, 264591, 55812038, 55811386, 85658542, 264760, 18108351, 18108359, 55811957, 265020, 265021, 33657023, 18108364, 55811576, 83373044, 18108385, 56526486, 264482

2140	87645655 (4279, 4280)	Novel Protein sim. GBank gij4417293jgb AAD20418  - (AC007019) unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264488, 264259, 29331824, 264104, 264108, 264509, 265006, 264759, 265018, 264448, 264288, 21908768, 55811957, 265021, 33657023, 27488265, 35698423, 264636, 264556, 264557, 264559, 264566
2141	79623986 (4281, 4282)		UNCLASSIFIED	265020, 264693
2142	80041222 (4283, 4284)		UNCLASSIFIED	263978
2143	84140051 (4285, 4286)	Novel Protein sim. GBank gij2135766jpij S53362 - mucin 5AC (clone JER47) - human (fragment)	UNCLASSIFIED	22278997, 29331827, 264907, 265020, 60432113
2144	94320114 (4287, 4288)	Novel Protein sim. GBank gij2076483 (U43200) - antifreeze glycoprotein AFGP polyprotein precursor [Boreogadus saida]	UNCLASSIFIED	65274572, 264259, 29331824, 29331827, 264906, 264908, 264591, 265011, 87168559, 264600, 265019, 264288, 264768, 21908765, 21906767, 55811576, 35698423, 65274791, 22279002
2145	20564305 (4289, 4290)		UNCLASSIFIED	263978
2146	87010515 (4291, 4292)	Novel Protein sim. GBank gij1255871 (U53341) - short region of weak similarity to bovine membrane receptor p63 (PIR: S28503) [Caenorhabditis elegans]	UNCLASSIFIED	264909, 60433356, 264686
2147	80432911 (4293, 4294)	Novel Protein sim. GBank gij3080338 emb CAA18718.1  - (AL022603) putative protein [Arabidopsis thaliana]	UNCLASSIFIED	264907, 264768, 264769, 18108385
2148	80048811 (4295, 4296)	Novel Protein sim. GBank gij728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII	UNCLASSIFIED	264593
2149	87362022 (4297, 4298)	Novel Protein sim. GBank gij119863 sp P20693 FCE2_MOUSE - LOW AFFINITY IMMUNOGLOBULIN EPSILON FC RECEPTOR (LYMPHOCYTE IGE RECEPTOR) (FC-EPSILON-RII) (CD23)	glycoprotein	29331824, 29331826, 35698052, 264758, 87168474, 265018, 52644150, 33657109
2150	84140059 (4299, 4300)	Novel Protein sim. GBank gij5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	22278998, 29331822, 29331824, 29331828, 264764, 264769, 21908766, 264488
2151	95353241 (4301, 4302)	Novel Protein sim. GBank gij5689407 dbj BAA82987.1  - (AB028958) KIAA1035 protein [Homo sapiens]		22278996, 56994075, 22278999, 60432049, 264259, 29331822, 29331824, 29331826, 35698052, 29331828, 264508, 264511, 60433356, 264758, 264598, 33109954, 60174639, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264288, 264689, 21906765, 21906766, 21908768, 265020, 60170615, 33657109, 33657182, 33657349, 18108370, 264635, 264557, 60170394, 18108385, 87168518, 22279000, 18108397, 18108398, 265007, 264591, 265011, 18108351, 18108368, 18108374, 18108388
2152	76321640 (4303, 4304)	Novel Protein sim. GBank gij3452473 (AF084205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]	kinase	

2153	88313371 (4305, 4306)	Novel Protein sim. GBank gi 4758704 ref NP_004216.1 pMASL - MFL-amplified sequences with leucine-rich tandem repeats 1	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264488, 263994, 52646842, 22278996, 22278998, 22278999, 264259, 29331822, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 56182435, 264510, 264511, 264512, 264758, 87188474, 87188559, 265017, 265019, 264760, 264288, 264369, 264788, 264887, 264769, 52644229, 21906766, 21906768, 35695917, 33657023, 33657109, 35895855, 264631, 264632, 264635, 264636, 264639, 18108385, 264483, 264564, 264488
2154	87408034 (4307, 4308)	Novel Protein sim. GBank gi 225150 prf 1209265U - chorion protein B11 [Bombyx mori]		UNCLASSIFIED	56994075, 264094, 265009, 265019, 264288, 21906767, 35695917
2155	87424072 (4309, 4310)			UNCLASSIFIED	18108392, 18108398, 22278996, 264259, 29331824, 265008, 265010, 265011, 265017, 265019, 264288, 264686, 265020, 264693, 264628, 56182323
2156	84295205 (4311, 4312)	Novel Protein sim. GBank gi 3970966 (AC004974) - spa-1 like; similar to AF026504 (PID:g2555183) [Homo sapiens]			265007, 264684
2157	87316344 (4313, 4314)	Novel Protein sim. GBank gi 1076211 prf S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii		UNCLASSIFIED	264591
2158	86444218 (4315, 4316)	Novel Protein sim. GBank gi 4650844 dbj BAA77027.1 - (AB026180) - Kelch motif containing protein [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	dna_rna_bind	29331822, 264112, 265009, 264691, 33657023, 264634
2159	80083729 (4317, 4318)	Novel Protein sim. GBank gi 2879925 dbj BAA24826 - (AB007897) KIAA0437 [Homo sapiens]			264634
2160	16283674 (4319, 4320)	Novel Protein sim. GBank gi 1504006 dbj BAA13202 - (D86968) simliarto human ZFY protein. [Homo sapiens]		UNCLASSIFIED	265008
2161	87739131 (4321, 4322)			UNCLASSIFIED	55274572, 264508, 264805, 264906, 264907, 264908, 52644045, 264809, 265007, 264810, 264591, 264592, 264593, 55812038, 264586, 264758, 265011, 264600, 264762, 264763, 264683, 264764, 264288, 264766, 264688, 264768, 264789, 264689, 265020, 264691, 264628, 264629, 263978, 264632, 264634, 264557, 264638, 264639, 18108385, 264563, 264566, 264567
2162	84319526 (4323, 4324)			UNCLASSIFIED	56182575, 22278996, 264093, 264683, 33657023, 65274620, 60432113
2163	95417158 (4325, 4326)	Novel Protein sim. GBank gi 3876537 emb CAA98270 - (Z73974) cDNA EST yk29115.3 comes from this gene; cDNA EST yk29115.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	
2164	80569456 (4327, 4328)		Contains protein domain (PF01006) - Hepatitis C virus non-structural protein NS4a	collagen	264603, 264637, 264565

2165	84329169 (4329, 4330)	Novel Protein sim. GBank gj1086784 (U41107) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264906, 29331830, 56182435, 265009, 21906754, 33657084, 265011, 265019, 264448, 264288, 264369, 21906765, 21906768, 21906769, 265020, 265021, 264691, 264692, 33657023, 65274620, 35695855, 264556, 60170394, 83373044, 60432113, 22279002, 264567
2166	87618934 (4331, 4332)	Novel Protein sim. GBank gj2706522[emb]CAA75816] - (Y15895) ubiquitin activating enzyme [Drosophila melanogaster]	ubiquitin		52645156, 22278994, 22278998, 66714117, 29331828, 52644045, 265018, 265019, 264369, 21906765, 21906767, 21906768, 21906769, 265021, 265022, 264693, 27486262, 35695763, 18108376, 56526486, 87168518, 264567
2167	87716864 (4333, 4334)	Novel Protein sim. GBank gj2224713[dbj]BAA20840] - (AB002364) KIAA0366 [Homo sapiens]		UNCLASSIFIED	56182575, 35696286, 29331824, 29331826, 29146498, 56182435, 265008, 265009, 264592, 264593, 33657402, 33109954, 265011, 265017, 265018, 18108351, 264369, 21906764, 21906765, 21906768, 29148627, 21906769, 52644150, 33657109, 35696423, 18108381, 18108384, 18108385, 60432113, 264567
2168	86999334 (4335, 4336)	Novel Protein sim. GBank gj4321407[gb]AAD15748] - (AF047690) ATP-binding cassette protein M-ABC1 [Homo sapiens]	Contains protein domain (PF00664) - ABC transporter transmembrane region.	transport	66714117, 29331827, 264907, 264511, 264591, 265018, 264764, 264683, 264766, 264768, 264568
2169	87886937 (4337, 4338)	Novel Protein sim. GBank gj5106521[gb]AAD39741.1[AF105365] K-CI cotransporter KCC4 [Homo sapiens]		UNCLASSIFIED	264629, 264555, 264559
2170	84141033 (4339, 4340)	Novel Protein sim. GBank gj1508521[gb]AAD39741.1[AF105365] K-CI cotransporter KCC4 [Homo sapiens]		UNCLASSIFIED	65274572, 56182575, 22278997, 22278998, 264259, 29331825, 264509, 264906, 56182435, 60433438, 55812038, 264596, 55811386, 265019, 264762, 264763, 264448, 264764, 264684, 264288, 264766, 264685, 56181562, 264689, 55811957, 265020, 264535, 264691, 33657109, 60431528, 18108374, 35696423, 55811576, 65274791, 264634, 264639, 264558, 87168518, 60432113, 264564
2171	80194050 (4341, 4342)			UNCLASSIFIED	264369, 265020, 264558
2172	85452460 (4343, 4344)			UNCLASSIFIED	264259, 264558
2173	87036740 (4345, 4346)	Novel Protein sim. GBank gj4309681[gb]AAD15478] - (AC006930) R33423_1 [Homo sapiens]		UNCLASSIFIED	264369
2174	95003286 (4347, 4348)	Novel Protein sim. GBank gj2493778[sp]Q09456[YQ35_CAEEL - PUTATIVE CUTICLE COLLAGEN C09G5.5			264906, 35695855, 264555, 264557

2175	94325850 (4349, 4350)	Novel Protein sim. GBank gi 1263287 (U47855) - fibroin-3 [Araneus diadematus]		UNCLASSIFIED	264488, 35696286, 20281099, 29331826, 60432289, 35696052, 264109, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265008, 264910, 264591, 264594, 264595, 264596, 264758, 55812038, 265011, 264600, 264603, 264760, 264762, 264448, 264764, 264288, 264766, 264686, 264687, 21906768, 55811957, 35695917, 265020, 265022, 264691, 264692, 33657023, 264693, 264628, 264629, 55811578, 35696423, 65274791, 35695855, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 264558, 18108385, 60432113, 264563, 264564, 264565, 264566, 264486, 264567
2176	88223392 (4351, 4352)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII	Contains protein domain (PF00805) - Pentapeptide repeats (8 copies)	oncogene	52644507, 52646842, 22278994, 35696286, 22278996, 22278999, 29331826, 29331827, 35696052, 29331828, 33656970, 29331830, 264910, 33657402, 264758, 52644296, 87168559, 265018, 264689, 21906765, 21906767, 21906768, 35695917, 52644150, 264690, 33657023, 33657109, 52645129, 33657182, 27486261, 27486262, 33657349, 18108376, 18108377, 35695855, 87168518, 60432113, 264404, 22279000, 264486
2177	94128942 (4353, 4354)	Novel Protein sim. GBank gi 5454072 ref NP_006416.1 pSLU7 - step II splicing factor SLU7		kinase	18108392, 22278997, 22278999, 264093, 33657402, 265019, 264448, 264766, 264689, 21906767, 21906768, 21906769, 265021, 33657023, 18108370, 18108374, 60432113, 22279002
2178	87601557 (4355, 4356)	Novel Protein sim. GBank gi 473407 (U08215) - NST-1 [Mus musculus]	Contains protein domain (PF00012) - Hsp70 protein	eph	264486, 22278996, 22278999, 29331824, 29331825, 29331826, 29331827, 52644296, 87168474, 18108370, 35695855, 22279002
2179	87316275 (4357, 4358)			UNCLASSIFIED	60424269, 264760, 264628, 264632



2180	95351397 (4358, 4360)	Novel Protein sim. GBank gij3122317[sp]P90848[KMHB_DICDI - MYOSIN HEAVY CHAIN KINASE B (MHCK B)]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	52644507, 22278994, 35696286, 22278997, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331828, 33656970, 264508, 264509, 264906, 264908, 29331830, 264909, 264510, 265006, 265007, 33657402, 55812038, 21906754, 87168474, 87168559, 265017, 265018, 265019, 264763, 264682, 264683, 264684, 264288, 264686, 21906765, 21906766, 21906768, 265020, 265021, 265022, 52644150, 33657023, 33657109, 27486265, 33657349, 18108374, 35696423, 35695855, 263981, 60170394, 18108385, 56526486, 87168518, 60432113, 22278000, 264482, 264566, 264567, 264486 29331827, 264369, 18108376, 264564
2181	85764930 (4361, 4362)	Novel Protein sim. GBank gij3024689[sp]Q15542[T2D4_HUMAN - TRANSCRIPTION INITIATION FACTOR TFIID 100 KD SUBUNIT (TAFII100)]		kinase	
2182	87637731 (4363, 4364)	Novel Protein sim. GBank gij5420387[emb]CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	22278996, 22278997, 22278999, 264259, 29331822, 56182435, 264112, 264764, 264288, 21906767, 21906768, 21906769, 33657109, 18108376, 60170394, 22278000, 22278002 264760
2183	85460649 (4365, 4366)	Novel Protein sim. GBank gij3873406[g]AAC77482.1] - (U17129) unknown [Rhodococcus erythropolis]		Inf	29331822, 29331825, 29331826, 56182435, 265011, 264685, 264686, 21906768, 18108370, 264629, 284631, 264638, 264557
2184	87760890 (4367, 4368)	Novel Protein sim. GBank gij3114713 (AF061346) - Ecp1 protein [Mus musculus]			29331824, 264907, 66712502, 264757, 265019, 264288, 264692, 56526486
2185	87826463 (4369, 4370)	Novel Protein sim. GBank gij5106950[g]AAD39908.1]AF11361 - (AF113615) FH1/FH2 domain-containing protein FHOS [Homo sapiens]		ATPase_associated	264259, 29331822, 29331824, 29331826, 56182435, 264592, 55812038, 264760, 264766, 55811957, 33657023, 55811576, 56182323, 264563
2186	87739227 (4371, 4372)	Novel Protein sim. GBank gij2864625[emb]CAA16972] - (AL021811) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	35696052, 264905, 284906, 264907, 264908, 264510, 264511, 265008, 264910, 264758, 265019, 264762, 264681, 264766, 264769, 35695917, 264692, 35696423, 264631, 264835, 264637, 18108388, 264566, 264486 18108398, 56994075, 264259, 29331824, 29331825, 66714117, 29331827, 264906, 29331830, 265018, 265020, 265021, 56182323, 264559, 22279000, 22279002
2187	87388173 (4373, 4374)			UNCLASSIFIED	
2188	87771708 (4375, 4376)	Novel Protein sim. GBank gij5107816[g]AAD40129.1]AF14941 - (AF149413) contains similarity to histone deacetylases; Pfam PF00850. Score=13.3, E=5e-10, N=1 [Arabidopsis thaliana]		histone	
2189	85693573 (4377, 4378)	Novel Protein sim. GBank gij3452357 (AF075724) - unknown [Legionella pneumophila]	Contains protein domain (PF01596) - O-methyltransferase		22278996, 264259, 29331826, 21906754, 264369, 264288, 263987

2190	87639197 (4379, 4380)	Novel Protein sim. GBank gi 132675 sp P29315 RIIN_RAT - RIBONUCLEASE INHIBITOR		nucleaseinhib	22278996, 22278999, 29331822, 29331824, 29331826, 265008, 264910, 60170831, 55812038, 52844296, 265010, 265018, 264685, 264688, 56181582, 21906769, 35695917, 265022, 60170394, 22279000
2191	95198928 (4381, 4382)	Novel Protein sim. GBank gi 5327002 emb CAB46272.1  - (Y18503) XAP-S-like protein [Homo sapiens]			29331825, 29331826, 29331830, 264510, 264511, 264910, 264593, 264594, 264556, 264559
2192	11126316 (4383, 4384)	Novel Protein sim. GBank gi 482800 sp P34400 MI10_CAEEL - MIG-10 PROTEIN	Contains protein domain (PF00169) - PH domain		264558
2193	94140073 (4385, 4386)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	56181686, 29331825, 29331827, 264508, 264909, 265008, 284592, 60432229, 264288, 264684, 264766, 35695917, 33657023, 60431802, 60431528, 55810764, 55811576, 65274791, 35695855, 60431850, 56182323, 60432113
2194	21418714 (4387, 4388)	Novel Protein sim. GBank gi 2773341 (AF040954) - putative protein phosphatase 1 nuclear targeting subunit [Rattus norvegicus]			264592
2195	88083023 (4389, 4390)	Novel Protein sim. GBank gi 2832763 emb CAA15685.1  - (AL009191) /prediction=(method); /prediction=(method); /match=(desc); /match=(desc); /motif=(desc); [Drosophila melanogaster]		UNCLASSIFIED	22278996, 22278999, 35696052, 265006, 21908754, 265017, 35695917, 265021, 265022, 35695855
2196	95091631 (4391, 4392)	Novel Protein sim. GBank gi 5262487 emb CAB45699.1  - (AL080076) hypothetical protein [Homo sapiens]		collagen	56182575, 35696286, 22278997, 22278999, 264259, 29331822, 66714117, 60432289, 29331827, 35696052, 29331828, 264508, 52844045, 56182435, 264510, 265007, 265008, 265009, 60433438, 55812038, 265010, 265011, 264448, 264288, 264686, 264687, 52844229, 21906765, 21906766, 21906767, 35695917, 265022, 264691, 33657023, 264693, 18108370, 18108376, 35696423, 55811576, 65274791, 35695855, 264636, 56182323, 18108385
2197	95073813 (4393, 4394)	Novel Protein sim. GBank gi 4928567 gb AAD34044.1 AF15180 - (AF151807) CGI-49 protein [Homo sapiens]			264768, 264769, 21906765, 21906766, 21906767, 29148627, 55811957, 35696286, 265020, 22278998, 265021, 264259, 33657023, 264693, 29331824, 35696052, 29331828, 18108370, 35695855, 264113, 265008, 264910, 60432229, 56182323, 33657402, 264758, 83373044, 21906754, 265018, 265019, 22279002, 264482, 264448, 264565, 264288, 264369
2198	88060914 (4395, 4396)	Novel Protein sim. GBank gi 3548787 (AC005622) - R30953.1 [Homo sapiens]		UNCLASSIFIED	

2199	86054355 (4397, 4398)	Novel Protein sim. GBank gij279372 (AC002505) - hypothetical protein [Arabidopsis thaliana]				264105, 264110, 264112, 264688, 55811957, 33657023, 264692, 263967, 20281071, 56526486
2200	87405385 (4398, 4400)	Novel Protein sim. GBank gij3043634[dbj BAA25481] - (AB011127) KIAA0555 protein [Homo sapiens]	struct			29331824, 264763, 264768
2201	94316872 (4401, 4402)	Novel Protein sim. GBank gij3913470[sp O57314 DHXB_ANAPL - PUTATIVE STEROID DEHYDROGENASE SPM2	Contains protein domain (PF00106) - short chain dehydrogenase			29331824, 35696052, 264905, 264907, 33657402, 55811386, 265017, 265018, 265019, 264288, 21906768, 35695917, 265020, 265022, 33657023, 33657109, 27486261, 18108370, 35696423, 35695855, 264555, 264556, 83373044, 87168518, 60432113
2202	91672385 (4403, 4404)	Novel Protein sim. GBank gij5262665[emb CAB45767.1] - (AL080186) hypothetical protein [Homo sapiens]	UNCLASSIFIED			264489, 264259, 29331824, 60432289, 35696052, 264905, 264909, 264592, 265017, 265018, 265019, 18108351, 264762, 264448, 264369, 264288, 264766, 21906765, 21906766, 264690, 264691, 264692, 33657109, 264634, 264636, 264555, 264639, 264558, 264559, 83373044, 18108385, 264404, 22279002, 264482
2203	87761832 (4405, 4406)	Novel Protein sim. GBank gij1172845[sp P48629 RB25_RABIT - RAS-RELATED PROTEIN RAB-25	Contains protein domain (PF00071) - Ras family	glycoprotein		52646365, 56994075, 264259, 29331822, 29331826, 29331827, 29331828, 264910, 265010, 265011, 87168558, 265018, 265019, 264605, 264288, 21906769, 35695917, 33657023, 264692, 33657109, 35695763, 18108376, 264638, 22279000, 264566, 264567
2204	88088671 (4407, 4408)	Novel Protein sim. GBank gij121036[sp P29348 GBT3_RAT GUANINE NUCLEOTIDE-BINDING PROTEIN G(T), ALPHA-3 SUBUNIT (GUSTOCIN ALPHA-3 CHAIN)	Contains protein domain (PF00503) - G-protein alpha subunit	UNCLASSIFIED		
2205	94147589 (4409, 4410)	Novel Protein sim. GBank gij4589480[dbj BAA76768.1] - (AB023141) KIAA0924 protein [Homo sapiens]	Contains protein domain (PF00096) - dna_ma_bind Zinc finger, C2H2 type			18108394, 18108397, 56182575, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 264906, 265007, 265008, 265009, 80432229, 265010, 265011, 265018, 264683, 264288, 264369, 264686, 21906766, 21906768, 21906769, 264690, 264691, 264693, 18108388, 55811576, 65274791, 264634, 18108381, 18108384, 60432113, 22279002, 264563, 264568, 264591
2206	20620008 (4411, 4412)	Novel Protein sim. GBank gij455753[ref NP_000372.1 pMID1 - midline 1 protein	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED		29331822, 56182181, 29331827, 35696052, 52644045, 265006, 265019, 56181562, 55811957, 265021, 33657023, 35695763, 35695855, 60170394, 60432113, 264568
2207	87787870 (4413, 4414)					
2208	86100830 (4415, 4416)	Novel Protein sim. GBank gij3986746 (AF105228) - tuftelin [Bos taurus]				264906, 265019, 18108351, 21906769
2209	87800420 (4417, 4418)		struct			264112, 265009, 264691, 18108385, 18108374, 264634, 20281166

2210	57152407 (4418, 4420)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		kinase	284603
2211	87341720 (4421, 4422)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		oncogene	264685, 264686, 18108365, 22278002, 264482
2212	91223924 (4423, 4424)	Novel Protein sim. GBank gi 3776027 emb CAA09214  - (AJ010475) RNA helicase [Arabidopsis thaliana]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	22278995, 22278997, 22278998, 264092, 264094, 28331822, 66714117, 29331826, 29331828, 264907, 52644045, 265009, 60170831, 21906754, 87168559, 265017, 265019, 18108351, 264683, 18108354, 264369, 264768, 264687, 52644229, 21906765, 21906766, 21906767, 21906768, 265021, 33657109, 18108370, 18108374, 264638, 56182323, 18108384, 18108387, 87168518, 264565
2213	91219309 (4425, 4426)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]			56182575, 22278996, 22278997, 35696052, 264905, 66712502, 264908, 264828, 56182435, 264112, 285008, 60431735, 60433438, 21906754, 265010, 265011, 265017, 265018, 265019, 18108351, 264765, 21906765, 21906768, 21906769, 265020, 265021, 264693, 264629, 263974, 263976, 18108379, 55811576, 264556, 264637, 264558, 83373044, 22278002, 264482, 264483

2214	95361453 (4427, 4428)	Novel Protein sim. GBank gil504325[refNP_000173.1]pHADH - hydroxyacyl- Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha su	Contains protein domain (PF00725) - 3-hydroxyacyl-CoA dehydrogenase	dehydrogenase	264488, 52644507, 18108394, 56182575, 22278994, 22278995, 35696286, 56994075, 22278997, 22278998, 22278999, 264490, 60432049, 264259, 52645080, 29331822, 29147620, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 20281100, 264509, 264907, 66712502, 264908, 29331830, 52644045, 56182435, 264510, 265006, 264511, 264512, 265007, 265008, 265009, 60170831, 264593, 60433356, 60433438, 33109954, 33657084, 52644298, 87168474, 265010, 265011, 87168559, 264601, 265017, 265018, 265019, 264687, 52644229, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 264532, 60170615, 264690, 52644150, 264691, 33657023, 264692, 18108364, 33657109, 33657182, 27486262, 27486264, 27486265, 35695763, 18108370, 264629, 60431528, 18108374, 18108376, 55810764, 35698423, 35695855, 264634, 264636, 52644332, 264638, 264558, 60170394, 18108381, 56182323, 83373044, 18108385, 18108387, 18108388, 56526486, 87168518, 60432113, 22279002, 264482, 264564, 264565, 264566, 264909, 265006, 264555, 264558, 87168518
2215	95419208 (4428, 4430)	Novel Protein sim. GBank gil1947160 (AF000288) - weak similarity to collagens; glycine- and proline-rich (Caenorhabditis elegans)			
2216	87614046 (4431, 4432)	Novel Protein sim. GBank gil1572802 (U70854) - similar to Enterococcus faecalis TRAB (GI:388268) [Caenorhabditis elegans]		UNCLASSIFIED	264693
2217	80589404 (4433, 4434)	Novel Protein sim. GBank gil5031707[refNP_005503.1]pGARP - glycoprotein A repetitions predominant	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264288, 33657109, 264556
2218	85518254 (4435, 4436)	Novel Protein sim. GBank gil387863[embjCAA88953] - (Z49128) similar to cAMP-dependant protein kinase; cDNA EST EMBL: T00719 comes from this gene; cDNA EST yk465d8.3 comes from this gene; cDNA EST yk465d8.5 comes from this gene; cDNA EST yk492f4.3 comes from this gene; cDNA EST y...	Contains protein domain (PF00069) - Eukaryotic protein kinase domain		35696423, 264563
2219	87614048 (4437, 4438)	Novel Protein sim. GBank gil1572802 (U70854) - similar to Enterococcus faecalis TRAB (GI:388268) [Caenorhabditis elegans]	Contains protein domain (PF01963) - TraB family		264682, 264683, 264688, 264689, 264693, 18108370, 18108376

2220	95354165 (4438, 4440)	Novel Protein sim. GBank gi 4507261 ref NP_003145.1 pSTAT - statherin			264488, 18108394, 18108395, 35696286, 264259, 264097, 60432289, 264509, 264905, 264906, 264907, 29331830, 264908, 264909, 264510, 264511, 265007, 264512, 264910, 265009, 264593, 264594, 60433356, 264595, 55812038, 264758, 85658542, 265010, 264601, 264603, 265019, 264605, 264760, 264762, 264448, 264764, 264368, 264766, 18108357, 264768, 264887, 18108358, 264769, 55811957, 264690, 264691, 33657023, 264692, 18108362, 18108368, 264628, 264629, 18108374, 263978, 264634, 264635, 264636, 264637, 264638, 18108385, 264483, 264566, 264486, 264567
2221	88060927 (4441, 4442)	Novel Protein sim. GBank gi 3549154 (AC005625) - R27328_1 [Homo sapiens]			
2222	84425892 (4443, 4444)			UNCLASSIFIED	264908, 265020, 35695855
2223	95091849 (4445, 4446)			UNCLASSIFIED	265010, 264685, 264690, 264693, 264628, 263974, 263976, 55811576, 264555, 264638, 83373044, 264483
2224	87388515 (4447, 4448)	Novel Protein sim. GBank gi 3876005 emb CAA84789  - (Z35719) cDNA EST EMBL:D67419 comes from this gene; cDNA EST EMBL:C13853 comes from this gene; cDNA EST EMBL:C11578 comes from this gene; cDNA EST yk234a7.3 comes from this gene; cDNA EST yk234a7.5 comes from this gene; cDNA ES...	Contains protein domain (PF01958) - Domain of unknown function	UNCLASSIFIED	264259, 264509, 56182435, 265006, 265008, 265009, 264757, 21908754, 18108351, 264693, 18108374, 18108385
2225	85749484 (4449, 4450)	Novel Protein sim. GBank gi 1255847 (U53338) - C05E11.1 gene product [Caenorhabditis elegans]	transport		22278994, 22278995, 22278999, 52644045, 264600, 265019, 21908765, 21908769
2226	86978953 (4451, 4452)	Novel Protein sim. GBank gi 4826524 emb CAB42852.1  - (AL049848) hypothetical protein [Homo sapiens]			264259, 29331822, 29331824, 29331825, 29331827, 264508, 264908, 265007, 264691, 264634, 264486
2227	87721135 (4453, 4454)			UNCLASSIFIED	22278999, 265006, 265008, 18108354, 29148629, 29148784, 27486261, 18108374, 264637, 18108384
2228	91227337 (4455, 4456)	Novel Protein sim. GBank gi 606976 (U16800) - ribonucleoprotein [Xenopus laevis]	dna_ma_bind		264091, 264092, 264094, 29331822, 29331825, 66714117, 264693, 263972, 264639, 83373044, 264563
2229	88060931 (4457, 4458)	Novel Protein sim. GBank gi 3549155 (AC005625) - R27328_2 [Homo sapiens]		UNCLASSIFIED	

2230	95342815 (4459, 4460)	Novel Protein sim. GBank gij226154[prf]1412350A - DNA polymerase [Human adenovirus type 2]		UNCLASSIFIED	284488, 264768, 52644507, 264769, 21906765, 21906766, 21906767, 21906769, 22278995, 35695917, 22278996, 22278997, 22278998, 22278999, 265021, 264259, 52645129, 29331827, 264508, 264509, 264907, 18108370, 18108374, 35696423, 35695855, 265007, 264910, 264555, 33657402, 21906754, 18108387, 265010, 265018, 265019, 264760, 264288, 264587, 264563
2231	88060937 (4461, 4462)	Novel Protein sim. GBank gij3549154 (AC005625) - R27328_1 [Homo sapiens]		UNCLASSIFIED	
2232	87762581 (4463, 4464)	Novel Protein sim. GBank gij5281316[gbjAAD41476.1]AF13312 - (AF133124) transcription factor IIC63 [Homo sapiens]		transcript factor	18108394, 56182575, 22278995, 35696286, 22278997, 22278999, 264259, 29331827, 35696052, 264807, 56182435, 265006, 265007, 265008, 264910, 264758, 55812038, 264603, 265018, 265019, 18108351, 264682, 264764, 264683, 264369, 264288, 264686, 264687, 264689, 21906765, 21906786, 21906767, 21906769, 29148629, 35695917, 264690, 52644150, 264691, 33657023, 264693, 18108370, 18108374, 55811576, 35695855, 264639, 18108385, 264564, 264906, 33657402, 265018, 264288, 264686, 265020, 264635, 18108385
2233	87755282 (4465, 4466)	Novel Protein sim. GBank gij4249733[gbjAAD13780] - (AF109377) IdBP [Mus musculus]		kinase	56182435, 264369, 264688, 21906765, 265020, 264693, 264556, 56526486
2234	87771817 (4467, 4468)	Novel Protein sim. GBank gij1706559[spP54352]EAS_DROME - ETHANOLAMINE KINASE (EASILY SHOCKED PROTEIN)		UNCLASSIFIED	22278997, 264563
2235	91012316 (4469, 4470)	Novel Protein sim. GBank gij4972734[gbjAAD34762.1] - (AF132174) unknown [Drosophila melanogaster]	Contains protein domain (PF00400) - WD domain, G-beta repeat		284569, 264687, 264769, 265022, 264259, 60432049, 264691, 29331826, 60432289, 20281149, 264906, 264907, 264511, 265008, 265009, 264634, 264635, 264636, 264555, 264556, 264557, 264558, 60433356, 264595, 264559, 60433438, 60432113, 264761, 264762, 264763, 264764
2236	88003131 (4471, 4472)	Novel Protein sim. GBank gij1082675[prf]B53814 - p20 protein - human	Contains protein domain (PF00011) - Hsp20/alpha crystallin family	eph	21906767, 21906768, 21906769, 22278995, 22278996, 22278997, 22278998, 265020, 265021, 22278999, 264259, 29331824, 29331826, 29331827, 35695763, 18108376, 35695855, 265007, 60432228, 33657402, 60433356, 60433438, 83373044, 18108385, 21906754, 18108387, 60432113, 22278900, 265019, 22279002, 264482, 18108351, 264288
2237	81012318 (4473, 4474)	Novel Protein sim. GBank gij4972734[gbjAAD34762.1] - (AF132174) unknown [Drosophila melanogaster]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	

2238	94998657 (4475, 4476)		Contains protein domain (PF00286) - Viral coat protein		264509, 264907, 264628, 264634, 264564
2239	87798688 (4477, 4478)				29331825, 265009, 264369, 33657109, 18108370, 18108374, 284557, 264559
2240	94121471 (4479, 4480)	Novel Protein sim. GBank gij2882311 (AF051240) - probable ubiquitin-conjugating enzyme E2 [Picea mariana]	Ubiquitin-conjugating enzyme Ubiquitin		264488, 65274572, 56182575, 35696286, 22278997, 22278999, 264259, 29331827, 35696052, 264508, 52644045, 56182435, 264511, 265007, 265008, 265009, 60433356, 60433438, 55812038, 21908754, 33657084, 55811386, 265018, 265019, 18108351, 264683, 264288, 264768, 264687, 264688, 264768, 21908765, 21908768, 21908769, 35695917, 265021, 265022, 60170615, 52644150, 33657023, 33657182, 33657349, 35695763, 18108370, 35696423, 35695855, 87168518, 22278000
2241	80091951 (4481, 4482)			UNCLASSIFIED	264693, 264629
2242	91228075 (4483, 4484)	Novel Protein sim. GBank gij2494312 [sp]P70541E2BG_RAT - TRANSLATION INITIATION FACTOR EIF-2B GAMMA SUBUNIT (EIF-2B GDP-GTP EXCHANGE FACTOR)		synthase	22278995, 22278996, 22278997, 22278998, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 264509, 265007, 265008, 264598, 21908754, 265010, 265011, 265017, 265018, 265019, 264448, 264369, 264288, 52644229, 21908765, 21908766, 21908767, 21908768, 21908769, 265020, 265021, 33657109, 27486282, 27486284, 18108374, 35695855, 264634, 264637, 56182323, 83373044, 56526486, 87168518, 264564
2243	78902026 (4485, 4486)			UNCLASSIFIED	265008
2244	85723527 (4487, 4488)	Novel Protein sim. GBank gij2281143 (AF016417) - Similar to BZIP transcription factor [Caenorhabditis elegans] Novel Protein sim. GBank gij470340 (U00043) - similar to beta-mannosyltransferase [Caenorhabditis elegans]		UNCLASSIFIED	264604
2245	95318545 (4489, 4490)		Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	52645156, 22278995, 22278996, 22278997, 22278999, 29331822, 29331824, 29331827, 264907, 264512, 60433438, 264758, 21908754, 265011, 264603, 264764, 264687, 21908767, 21908768, 21908769, 55811957, 265022, 264691, 264629, 35696423, 264638, 18108387, 60432113, 22279000, 22279002, 264566



2246	94848710 (4481, 4492)	Novel Protein sim. GBank gi 4980896 dbj BAA78326.1  - (AB028069) activator of S phase Kinase [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	65274572, 22278995, 35696286, 22278996, 22278997, 22278999, 264259, 35696052, 264106, 264905, 264907, 265006, 265007, 265008, 60433438, 33109954, 87168559, 265018, 265019, 264288, 21906765, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265022, 27486264, 18108370, 18108374, 65274791, 35695855, 60432113
2247	87862542 (4483, 4494)	Novel Protein sim. GBank gi 854085 emb CAA58337  - (X83413) U88 [Human herpesvirus 6]	UNCLASSIFIED	UNCLASSIFIED	52645156, 52646365, 52645080, 35696052, 33656970, 52648317, 33657084, 265017, 21906768, 21906769, 35695917, 33657109, 52645129, 33657182, 27486261, 27486262, 33657349, 27486265, 18108387
2248	95412986 (4495, 4496)	Novel Protein sim. GBank gi 4758502 ref NP_004123.1 pHABP - hyaluronan-binding protein 2	Contains protein domain (PF00089) - Trypsin	calhepsin	264488, 264259, 264907, 29331830, 264909, 265007, 265009, 264595, 21906754, 65274444, 264603, 265019, 264762, 264448, 264288, 264689, 21906766, 55811957, 265021, 264691, 18108374, 264634, 264635, 264636, 264555, 264638, 264557, 264558, 264559, 18108383, 83373044, 18108385, 264486
2249	94685662 (4497, 4498)	Novel Protein sim. GBank gi 403846  (AF107772) - TcST11 [Trypanosoma cruzi]	Contains protein domain (PF00515) - TPR Domain	eph	264766, 264628, 264636, 264637
2250	79827508 (4499, 4500)	Novel Protein sim. GBank gi 3738140 emb CAA21241  - (AL031852) valyl-tRNA synthetase, mitochondrial precursor [Schizosaccharomyces pombe]	UNCLASSIFIED	UNCLASSIFIED	264908, 18108374
2251	87385863 (4501, 4502)	Novel Protein sim. GBank gi 3218467 emb CAA07090.1  - (AJ006529) putative phosphatase [Gallus gallus]	UNCLASSIFIED	UNCLASSIFIED	264259, 35696052, 264508, 56182435, 265009, 264592, 264593, 264780, 264448, 264684, 264288, 264690, 264628, 55811576, 264555, 264556, 264557, 264558, 264559, 264568
2252	87735867 (4503, 4504)	Novel Protein sim. GBank gi 4929325 gb AAD33953.1 AF14531 - (AF145316) vacuolar proton pump delta polypeptide [Homo sapiens]	Contains protein domain (PF01813) - ATP synthase subunit D	synthase	264092, 264094, 264259, 29331822, 66714117, 29331828, 264102, 264103, 264104, 264105, 264109, 264112, 264511, 265007, 60433356, 265010, 18108351, 21906767, 21906768, 264691, 263974, 263977, 264486, 264567
2253	91010703 (4505, 4506)		UNCLASSIFIED	UNCLASSIFIED	65274572, 265019

2254	95320031 (4507, 4508)	Novel Protein sim. GBank gi4502847 ref NP_001271.1 pCIRB - cold inducible RNA- binding protein	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	264569, 18108394, 18108398, 56182575, 56994075, 35696286, 22278999, 264094, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 35696052, 264106, 264508, 264509, 264906, 264907, 29331830, 66712502, 264908, 264909, 264510, 265008, 264511, 265007, 265008, 265009, 60170831, 60432229, 60433358, 60433438, 264758, 85658542, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264764, 264288, 264369, 264768, 264686, 264768, 264769, 21906765, 21906767, 55811957, 264691, 33657023, 264692, 18108362, 65274620, 263969, 264628, 18108370, 60431528, 263972, 264629, 18108372, 18108377, 18108379, 55811576, 35696423, 35695855, 264630, 264634, 264635, 264636, 264558, 263981, 264638, 56182323, 60170394, 264558, 18108381, 18108382, 83373044, 18108385, 87188518, 60432113, 22279002, 264482, 264584, 264585, 264486, 264567, 18108391
2255	81010548 (4508, 4510)	Novel Protein sim. GBank gi45541865 emb CAB51072.1  - (AL098858) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	struct	65274572, 56182575, 22278997, 22278999, 264259, 29331822, 29331825, 29331826, 29331827, 29331828, 264508, 264905, 264908, 264907, 68712502, 264908, 58182435, 264510, 264511, 265008, 264583, 264595, 21906754, 33109954, 87168474, 265011, 265017, 265018, 264682, 264764, 264369, 264288, 264768, 264685, 264686, 264768, 21906765, 21906766, 21906768, 21906769, 265020, 60170615, 52644150, 264690, 264692, 264693, 33657109, 33657349, 264632, 264636, 52644332, 56182323, 22279000, 22279002
2256	87020531 (4511, 4512)	Novel Protein sim. GBank gi3327174 dbj BAA31655  - (AB014580) KIAA0680 protein [Homo sapiens]		UNCLASSIFIED	264768, 264689, 18108374
2257	80088235 (4513, 4514)			UNCLASSIFIED	22278996, 22278999, 264681, 21906765, 21906768, 264567
2258	86090516 (4515, 4516)	Novel Protein sim. GBank gi3025446 (AC004528) - R32184_2 [Homo sapiens]	Contains protein domain (PF00060) - Ligand-gated ion channel	misc_channel	264908, 264592, 264764

2259	85364155 (4517, 4518)	Novel Protein sim. GBank gi 4884140 emb CAB43278.1  - (AL050110) hypothetical protein [Homo sapiens]		UNCLASSIFIED	18108396, 65274572, 56182575, 22278997, 22278998, 264259, 29331822, 29331827, 264905, 66712502, 264908, 264909, 56182435, 265007, 265008, 60432229, 33657084, 87168559, 18108351, 264448, 264883, 264288, 264369, 56181562, 265021, 60170815, 264690, 33657109, 60431528, 18108374, 52644332, 56182323, 18108385, 22279000, 22279002, 264482
2260	88084118 (4519, 4520)	Novel Protein sim. GBank gi 3080663 (AC004614) - similar to f-spondin proteins AB008086 (PID:g2529225) [Homo sapiens]	Contains protein domain (PF000090) - Thrombospondin type 1 domain	oxidase	56182575, 265020, 264905, 264906, 264908, 35696423, 264511, 264635, 55812038, 264758, 265018, 265019, 264605, 264760, 264563
2261	88074157 (4521, 4522)	Novel Protein sim. GBank gi 3334526 emb CAA16138  - (AL021306) predicted using FGENEH [Homo sapiens]		UNCLASSIFIED	
2262	81639292 (4523, 4524)	Novel Protein sim. GBank gi 4877759 gb AAD31421.1 AF12444 - (AF124440) MAGE tumor antigen D1 [Homo sapiens]			56182575, 22278998, 29331822, 29331825, 60432289, 29331827, 35696052, 264508, 66712502, 52644045, 56182435, 265006, 265008, 265009, 60433356, 55812038, 265010, 265017, 265019, 264288, 264369, 21906765, 21906767, 55811857, 35695917, 52644150, 33657023, 33657109, 55811576, 65274791, 56182323
2263	87602485 (4525, 4526)	Novel Protein sim. GBank gi 33411687 (AC003672) - hypothetical protein [Arabidopsis thaliana]			22278994, 22278997, 264907, 264828, 52644150, 18108361, 264893, 18108374
2264	87756525 (4527, 4528)	Novel Protein sim. GBank gi 1657601 (U66220) - unknown [Nannocystis exedens]		UNCLASSIFIED	264686, 264488, 264768, 264769, 264691, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 35695855, 264510, 264511, 264512, 265007, 265009, 264638, 264639, 264757, 264758, 18108385, 265011, 264760, 264564, 264565, 264764, 264566, 264486, 264766
2265	86918663 (4529, 4530)	Novel Protein sim. GBank gi 477072 pir A48018 - mucin 7 precursor, salivary - human	Contains protein domain (PF000096) - Zinc finger, C2H2 type	UNCLASSIFIED	264689, 264910, 264764
2266	87773458 (4531, 4532)	Novel Protein sim. GBank gi 3150479 (AF067212) - partial CDS [Caenorhabditis elegans]	Contains protein domain (PF01305) - Ribosomal protein L15 amino terminal region	ribosomal prot	22278995, 22278997, 22278999, 264259, 265006, 265007, 265009, 60433438, 21906754, 265010, 265011, 265017, 264448, 264883, 264288, 264689, 21906765, 21906768, 35695917, 265021, 18108374, 264638, 22279000, 22279002, 264566, 264487

2267	87395838 (4533, 4534)	Novel Protein sim. GBank gi 3560229 emb CAA20697.1  - (AL031530) hypothetical protein [Schizosaccharomyces pombe]		UNCLASSIFIED	35696286, 264259, 29331824, 29331825, 35696052, 29331828, 264905, 284509, 264907, 264908, 264909, 264512, 265009, 264910, 264593, 33657402, 265010, 265018, 264762, 264448, 264288, 264369, 264768, 52644229, 35695917, 264691, 33657023, 18108382, 33657108, 35698423, 264634, 18108381, 87168518, 264568
2268	85693867 (4535, 4536)	Novel Protein sim. GBank gi 728832 sp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII		cadherin	264488, 264259, 284509, 264595, 265010, 265017, 264766, 18108385, 264488
2269	86177877 (4537, 4538)	Novel Protein sim. GBank gi 103418 pir S17885 - Tcd37 protein - fruit fly (Drosophila melanogaster)		UNCLASSIFIED	56182575, 60432049, 265007, 265009, 264591, 87168559, 264605, 18108351, 21906764, 265020, 264628, 60431528, 264638, 18108385, 18108387, 60432113, 264763
2270	80410327 (4539, 4540)			Cyfo450	264909, 56182435, 265008, 55812038, 55811957, 33657023, 284693, 33657109, 55810764, 55811576, 56182323
2271	91010392 (4541, 4542)			UNCLASSIFIED	264905, 264908
2272	84208220 (4543, 4544)	Novel Protein sim. GBank gi 4176370 (AC005058) - similar to calcium-independent phospholipase A2; similar to AC004392 (PID:g3367519) [Homo sapiens]	Contains protein domain (PF00462) - Glutaredoxin		52645156, 22278996, 22278999, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 265006, 264593, 60433438, 21906754, 265018, 284689, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486264, 18108376, 35696423, 35695855, 264630, 52644332, 264558, 56182323, 22279002
2273	95014271 (4545, 4546)			histone linker histone H1 and H5 family	52645156, 22278997, 22278999, 52645080, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264908, 52644045, 264511, 265008, 265009, 60170831, 264591, 21906754, 33109954, 265011, 265018, 18108351, 264448, 264288, 264684, 264766, 21906765, 21906766, 21906767, 21906768, 52644150, 264693, 18108384, 35695763, 18108374, 35696423, 264634, 264557, 264638, 52644332, 83373044, 18108385, 56526486, 87168518, 22279002
2274	91640217 (4547, 4548)	Novel Protein sim. GBank gi 1480112 emb CAA67961  - (X99642) HP1-BP74 protein [Mus musculus]	Contains protein domain (PF00538) - linker histone H1 and H5 family		
2275	88082501 (4549, 4550)	Novel Protein sim. GBank gi 3165406 (AC004755) - fos37502.2 [Homo sapiens]	Contains protein domain (PF00122) - E1-E2 ATPase	transport	
2276	11287447 (4551, 4552)			UNCLASSIFIED	264555, 264556

2277	88084123 (4553, 4554)	Novel Protein sim. GBank gii2880079 (AC004142) - similar to murine leucine-rich repeat protein; possible role in neural development by protein-protein interactions; 93% similarity to D49802 (PID:g1369906) [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	22278999, 35696052, 265008, 265019, 264369, 265020, 265022, 55810764, 264404, 22279002
2278	94133079 (4555, 4556)	Novel Protein sim. GBank gii2618702 (AC002510) - unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	18108394, 22278997, 29331826, 60433356, 60433438, 21906754, 265018, 33657023, 264639, 83373044, 264565
2279	80419375 (4557, 4558)	Novel Protein sim. GBank gii119714spj13983JEXTN_TOBAC - EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)		UNCLASSIFIED	264766, 264565
2280	94238723 (4558, 4560)				
2281	95293048 (4561, 4562)	Novel Protein sim. GBank gii4240299[dbj]BAA74928.1] - (AB020712) KIAA0905 protein [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	transport	265008, 33109954, 265010, 265019, 265020, 264092, 264259, 29331822, 29331824, 29331826, 35696052, 264107, 264906, 264909, 52844045, 265006, 33657402, 60433356, 264758, 265011, 265019, 264681, 264683, 264684, 264688, 21906765, 21906767, 21906768, 21906769, 60170615, 264690, 52844150, 18108362, 264692, 18108368, 18108374, 263978, 264631, 18108381, 264559, 18108385, 56528486, 22279000, 264566, 264567
2282	87602829 (4563, 4564)	Novel Protein sim. GBank gii1537070 (U63840) - nucleoporin p54 [Rattus norvegicus]		UNCLASSIFIED	264488, 264259, 29331822, 29331824, 29331827, 29331828, 29331830, 33657402, 60433438, 87168474, 265019, 18108351, 21906767, 21906769, 55811957, 33657023, 52845129, 33657109, 33657182, 27486282, 263972, 55811576, 87168518, 20281169
2283	95362386 (4565, 4566)	Novel Protein sim. GBank gii2495729[sp]Q92556[Y281_HUMAN - HYPOTHETICAL PROTEIN KIA0281 (HA6725)]		UNCLASSIFIED	60424179, 56182575, 22278994, 35696286, 22278997, 22278999, 29331822, 29331824, 56182181, 29331825, 29331827, 35696052, 29146499, 264905, 66712502, 264908, 265007, 265009, 60432229, 264593, 60431735, 60433356, 33109954, 33657084, 55811386, 87168474, 265010, 265011, 265018, 265019, 55811150, 264683, 264369, 264288, 264688, 21906765, 21906767, 21906768, 28148627, 21908769, 55811957, 265020, 265022, 33657182, 27486261, 18108370, 264628, 18108374, 55810764, 18108379, 55811576, 35696423, 35695855, 264630, 60431850, 263991, 18108382, 83373044, 18108385, 18108387, 60432113, 22279000, 264482, 264567

2284	95414955 (4567, 4568)	Novel Protein sim. GBank gi 2498797 sp Q64311 PNAD_MOUSE - PROTEIN N- TERMINAL ASPARAGINE AMIDOHYDROLASE (PROTEIN NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN- AMIDASE) (PNAD) (PROTEIN NH2-TERMINAL ASPARAGINE AMIDOHYDROLASE) (PNA4)			60424179, 52644507, 18108394, 52646842, 22278994, 35696286, 22278996, 22278997, 22278999, 264259, 60432049, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331830, 52644045, 56182435, 33657402, 60433438, 33109954, 21906754, 85658542, 87188559, 265018, 265019, 55811150, 264682, 264369, 21906765, 21906768, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 33657023, 33657182, 27486262, 27486264, 27486265, 18108376, 55810764, 35696423, 35695855, 60431850, 87188518, 60432113, 264482, 264564
2285	87781484 (4569, 4570)	Novel Protein sim. GBank gi 3342234 (U93909) - nuclear antigen EBNA-1 [Cercopithecine herpesvirus 15]		collagen	35696052, 264905, 264907, 264908, 264909, 264512, 265009, 264910, 264595, 264760, 18108351, 264682, 264763, 264685, 264766, 264686, 264768, 264693, 264629, 35695855, 264631, 264634
2286	87737825 (4571, 4572)	Novel Protein sim. GBank gi 3873414 (U00043) - similar to D. melanogaster trithorax protein [Caenorhabditis elegans]		kinase	35696286, 56182435, 60170831, 264591, 60432229, 264592, 264593, 264594, 264595, 55812038, 264596, 87188474, 35695917, 264692, 55811576, 264555, 264557
2287	82986696 (4573, 4574)	Novel Protein sim. GBank gi 630905 pir S42731 - collagen alpha 1 chain - sea urchin (Hemicentrotus pulcherrimus) (fragment)	Contains protein domain (PF01391) - Collagen triple helix repeat (20 copies)	UNCLASSIFIED	264682
2288	94133083 (4575, 4576)	Novel Protein sim. GBank gi 728832 sp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII		tm7	22278995, 56994075, 22278997, 22278999, 264259, 60432289, 264508, 264512, 265008, 33657402, 265017, 265018, 265019, 18108351, 264446, 21906765, 21906766, 21906767, 21906768, 33657023, 264557, 22279000, 22279002
2289	88084133 (4577, 4578)	Novel Protein sim. GBank gi 2887497 (AC004144) - R34001_1 [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	
2290	88084137 (4579, 4580)	Novel Protein sim. GBank gi 2887497 (AC004144) - R34001_1 [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	264683
2291	84295281 (4581, 4582)	Novel Protein sim. GBank gi 3253120 (AC005175) - R31449_3 [Homo sapiens]		struct	18108394, 264907, 265006, 265009, 33108954, 52646317, 265010, 18108351, 264681, 264686, 264692, 18108370, 18108374, 18108385

2282	94328834 (4583, 4584)	Novel Protein sim. GBank gi 4803672 emb CAB42643.1  - (AJ133769) nuclear transport receptor [Homo sapiens]		UNCLASSIFIED	56182575, 35686286, 56994075, 29331824, 29331825, 35686052, 56182435, 60433438, 55812038, 33109954, 87168474, 87168559, 265018, 18108351, 264763, 264448, 264369, 264288, 56181562, 264769, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 33657023, 264693, 65274620, 33657109, 27486264, 264629, 55810764, 55811576, 35695855, 56182323, 56526486, 87168518, 22279000, 264567
2283	87759213 (4585, 4586)	Novel Protein sim. GBank gi 3252981 (AF068921) - Ras-binding protein SUR-9 [Mus musculus]	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	264488, 18108397, 35686286, 264092, 264259, 29331822, 29331826, 264906, 264908, 264511, 264512, 265009, 264910, 18108351, 264764, 264369, 264288, 264685, 264766, 265020, 265022, 264534, 35696423, 264631, 264637, 18108381, 56182323, 264639, 18108385, 264404, 264563, 264565
2284	86693580 (4587, 4588)	Novel Protein sim. GBank gi 2062680 (U88964) - HEM45 [Homo sapiens]	Contains protein domain (PF00929) - Exonuclease	nuclease	22278997, 22278998, 22278999, 264259, 29331822, 60432289, 29331828, 35696052, 265018, 264684, 264288, 264686, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 264692, 33657109, 18108376, 35696423, 35695855, 264634, 22279000, 22278002, 264563, 264486
2285	95312200 (4589, 4590)			UNCLASSIFIED	22278996, 60432289, 264682, 264683, 264689, 18108374
2286	80030781 (4591, 4592)				263974, 263978
2287	94321251 (4593, 4594)	Novel Protein sim. GBank gi 5689501 dbj BAA83034.1  - (AB029005) KIAA1082 protein [Homo sapiens]		transcriptfactor	264488, 65274572, 56182575, 22278997, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 35696052, 264907, 264908, 52644045, 56182435, 264112, 265006, 265007, 264910, 265009, 60433356, 33657402, 264595, 55812038, 21906754, 265011, 265018, 265019, 264448, 264764, 264288, 264766, 21906765, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 264690, 264691, 33657023, 264692, 264693, 33657109, 55811576, 56182323, 60170394, 83373044, 18108385, 56526486, 264564, 264486

2298	95312207 (4595, 4596)	Novel Protein sim. GBank gi 3875051 emb CAB02849  - (Z81050) predicted using Genefinder; similar to collagen; cDNA EST EMBL:D65564 comes from this gene; cDNA EST EMBL:D69046 comes from this gene; cDNA EST yk366b12.3 comes from this gene; cDNA EST yk366b12.5 comes from this gene ...	collagen	60424178, 56181686, 22278995, 35696286, 22278996, 22278998, 22278999, 264480, 264259, 28331822, 28331824, 66714117, 60424269, 35696052, 29331828, 66712502, 56182435, 264510, 265006, 60433438, 21906754, 33109954, 55811386, 265010, 265018, 55811150, 264762, 18108351, 264682, 264683, 264288, 264684, 264686, 264688, 56181562, 264689, 21906768, 21906787, 29148629, 55811957, 29148784, 35695917, 265020, 18108362, 33657023, 18108364, 33657109, 60431602, 18108370, 60431528, 18108374, 55810764, 35698423, 35695855, 264630, 264634, 60431850, 18108380, 56182323, 264558, 83373044, 18108385, 60432113, 22279000, 264482, 264567, 264486
2299	80193720 (4597, 4598)		UNCLASSIFIED	264369
2300	84124346 (4599, 4600)	Novel Protein sim. GBank gi 2443886 (AC002294) - Unknown protein [Arabidopsis thaliana]		264488, 22278996, 22278999, 264259, 29331824, 66714117, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 265008, 264910, 265009, 264758, 265010, 87188559, 264600, 265018, 264760, 264762, 18108351, 264764, 264766, 264768, 264769, 21906765, 21908767, 35695917, 265021, 264691, 33657023, 35695763, 18108370, 18108374, 35696423, 35695855, 264631, 264636, 264638, 18108385, 22279002, 264563
2301	91235725 (4601, 4602)	Novel Protein sim. GBank gi 2143637 pir j184505 - calcium-dependent actin-binding protein - rat	struct	264908, 264758, 265017, 21906765, 83373044, 264563
2302	88084141 (4603, 4604)	Novel Protein sim. GBank gi 2887497 (AC004144) - R34001_1 [Homo sapiens]	UNCLASSIFIED	52644045, 265019, 264288, 33657023, 18108370, 18108385
2303	94141439 (4605, 4606)	Novel Protein sim. GBank gi 4864194 emb CAB43220.1  - (AL049946) hypothetical protein [Homo sapiens]	Contains protein domain (PF000047) - struct Immunoglobulin domain	264259, 60432049, 264907, 264909, 264910, 60432229, 33657402, 265011, 265018, 264762, 264448, 264769, 264637, 264638, 83373044, 264486
2304	94840434 (4607, 4608)	Novel Protein sim. GBank gi 2494162 sp Q10005 YRY1_CAEEL - HYPOTHETICAL 39.8 KD PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR	UNCLASSIFIED	264259, 29331824, 21906767, 33657182, 33657349
2305	90635911 (4609, 4610)	Novel Protein sim. GBank gi 4972686 gb AA034738.1  - (AF132150) unknown [Drosophila melanogaster]		65274572, 22278996, 264908, 265006, 21906769, 264691, 264486



2306	95334940 (4611, 4612)	Novel Protein sim. GBank gi 4929565 gb AAD34043.1 AF151806 - (AF151806) CGI-48 protein [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinasereceptor	264488, 22278995, 22278996, 35695286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35698052, 284508, 264905, 264907, 29331830, 264908, 264909, 264510, 265008, 264511, 265007, 264512, 265008, 265009, 33657402, 21906754, 8568542, 265010, 265011, 264601, 265017, 265018, 264604, 265019, 18108351, 264448, 264288, 264766, 264768, 21906765, 21906766, 21906767, 21906768, 29148629, 29148784, 35695917, 265020, 265021, 265022, 33657023, 264692, 18108370, 18108374, 18108376, 35696423, 35695855, 264630, 264634, 264635, 264636, 264637, 264638, 264639, 18108382, 18108385, 18108387, 264563, 264566, 264486 264828
2307	79415283 (4613, 4614)	Novel Protein sim. GBank gi 4758732 ref NP_004522.1 pMOCS - molybdenum cofactor synthesis 2		UNCLASSIFIED	264828
2308	87608409 (4615, 4616)	Novel Protein sim. GBank gi 4758732 ref NP_004522.1 pMOCS - molybdenum cofactor synthesis 2		synthase	35695286, 264259, 29331822, 29331824, 264112, 264512, 264757, 21906754, 264288, 264690, 27486264, 264631, 264634, 264404 18108397, 22278996, 22278997, 22278998, 22278999, 60432049, 29331822, 29331826, 60432289, 66712502, 60432228, 60433356, 60433438, 65274444, 265010, 264600, 264681, 264448, 264683, 264288, 21906766, 21906768, 265020, 264691, 264692, 264693, 65274620, 65274791 264508
2309	95337218 (4617, 4618)	Novel Protein sim. GBank gi 3878058 emb CAB17070  - (Z99942) cDNA EST EMBL:D73444 comes from this gene; cDNA EST EMBL:D70905 comes from this gene; cDNA EST EMBL:D72208 comes from this gene; cDNA EST EMBL:D75030 comes from this gene; cDNA EST EMBL:D72944 comes from this gene; cDN...		glycoprotein	22278999, 60432049, 29331822, 29331826, 60432289, 66712502, 60432228, 60433356, 60433438, 65274444, 265010, 264600, 264681, 264448, 264683, 264288, 21906766, 21906768, 265020, 264691, 264692, 264693, 65274620, 65274791 264508
2310	79601688 (4618, 4620)	Novel Protein sim. GBank gi 2137337 pir J148281 - gene mCBP protein - mouse	Contains protein domain (PF00013) - KH domain	UNCLASSIFIED	264508
2311	87721189 (4621, 4622)	Novel Protein sim. GBank gi 2137337 pir J148281 - gene mCBP protein - mouse		transcriptfactor	18108397, 56182575, 22278996, 56994075, 264259, 29331824, 29331827, 264508, 264907, 56182435, 264510, 264511, 265006, 264512, 265007, 265008, 265009, 60433438, 33109954, 265010, 265011, 264603, 265017, 18108351, 264762, 264683, 264288, 264369, 264888, 33657023, 20281149, 20281069, 264628, 263972, 55811576, 35696423, 20281071, 264632, 264636, 18108385, 18108387, 87168518, 22279000, 264563, 264486

2312	87549681 (4623, 4624)	Novel Protein sim. GBank gj 2911264 (AC002550) - Unknown gene product [Homo sapiens]			56182575, 56994075, 35696286, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 52644045, 60432229, 60433356, 55812038, 33109954, 21906754, 87168474, 265018, 18108351, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 265022, 60170615, 33657023, 27486261, 27486264, 35696423, 35695855, 18108385, 22278000, 22278002 263981
2313	80042533 (4625, 4626)	Novel Protein sim. GBank gj 3043626 dbj BAA25477  - (AB011123) KIAA0551 protein [Homo sapiens]			
2314	94313401 (4627, 4628)	Novel Protein sim. GBank gj 5596714 emb CAB51401.1  - (AL035398) dJ796117.2 (CGI-51) [Homo sapiens]		UNCLASSIFIED	52644507, 52646365, 52646842, 22278996, 22278997, 22278999, 264259, 52645080, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 52644045, 265008, 264593, 60433356, 60433438, 264758, 33109954, 265010, 265017, 265018, 265019, 284288, 264369, 21906765, 21906766, 21906768, 35695917, 52644150, 33657023, 33657109, 52645129, 33657349, 35695763, 18108374, 35696423, 35695855, 52644332, 22279000, 22278002, 264563, 264567
2315	80430119 (4629, 4630)			UNCLASSIFIED	264905, 264906, 264767, 264768, 264693, 55811576, 264635, 56182323, 18108385
2316	94312191 (4631, 4632)	Novel Protein sim. GBank gj 5531827 gb AAD44488.1  - (AF078856) p47 [Homo sapiens]	Contains protein domain (PF00789) - UBX domain		52644507, 52645158, 52646365, 22278994, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 52645080, 29331824, 29331825, 29331826, 29331827, 35696052, 29331828, 264906, 264907, 29331830, 52644045, 56182435, 264511, 265007, 265008, 265009, 60170831, 60433438, 21906754, 52646317, 33109954, 33657084, 52644296, 87168474, 265010, 87168559, 265017, 265018, 265019, 264681, 264763, 264448, 264683, 264369, 52644229, 21906764, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 52644150, 33657023, 18108362, 52645129, 33657182, 33657349, 35695763, 18108370, 18108376, 35696423, 35695855, 264631, 264556, 52644332, 83373044, 18108385, 18108387, 87168518, 60432113, 22279000, 264566, 264567

2317	87020571 (4633, 4634)			UNCLASSIFIED	22278998, 60432049, 264810, 60432229, 264686, 264687, 264688, 264689, 264558, 18108385
2318	79959879 (4635, 4636)			UNCLASSIFIED	265006, 264910
2319	95101781 (4637, 4638)	Novel Protein sim. GBank gi 5262613 emb CAB45746.1  - (AL080155) hypothetical protein [Homo sapiens]			264488, 264569, 18108396, 52646365, 22278994, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 264259, 52645080, 29331825, 29331826, 29331827, 29331828, 29331830, 56182435, 60170831, 60432229, 60431735, 33657402, 21906754, 52644298, 87168474, 265011, 87168559, 265017, 265018, 265019, 18108351, 264448, 18108354, 264288, 264369, 52844229, 21906764, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 265022, 52644150, 33657023, 52645129, 33657109, 27486264, 33657349, 35695763, 18108370, 18108376, 18108379, 35696423, 264558, 83373044, 18108385, 56526486, 87168518, 264564, 264565, 264566
2320	91622426 (4638, 4640)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - III  ALU SUBFAMILY SQ WARNING ENTRY III		kinase	22278994, 60432049, 60432289, 29331827, 264511, 265008, 52646317, 265017, 265018, 21906765, 18108372, 18108387, 22279002
2321	94320377 (4641, 4642)	Novel Protein sim. GBank gi 3873837 emb CAB02700  - (Z81029) Similarity to S.pombe hypothetical protein C1D4.09C (SW:Q10154); cDNA EST EMBL: T00543 comes from this gene; cDNA EST EMBL: T01082 comes from this gene; cDNA EST EMBL: T01321 comes from this gene; cDNA EST EMBL: T02288 com...		UNCLASSIFIED	264488, 264687, 18108394, 264689, 21906765, 18108397, 18108398, 21906767, 21906768, 65274791, 22278995, 35695855, 22278998, 265021, 265022, 264510, 265006, 264511, 264512, 265008, 60170615, 264555, 264636, 264556, 18108361, 264259, 60432229, 33657023, 264557, 264558, 264693, 60433356, 264559, 60433438, 29331824, 18108365, 18108348, 18108384, 29331825, 18108385, 33109954, 29331827, 56526486, 29146499, 265011, 60432113, 265017, 265018, 264508, 264563, 264482, 264509, 18108351, 264448, 264907, 264682, 18108370, 264683, 264908, 264288, 264909, 18108354, 264486, 264567
2322	87803165 (4643, 4644)	Novel Protein sim. GBank gi 5678957 emb CAB51685.1  - (AL109630) BACR7A4.y [Drosophila melanogaster]	Contains protein domain (PF00106) - short chain dehydrogenase	dehydrogenase	22278996, 264907, 264511, 264757, 18108351, 264788, 264638

2323	94840445 (4645, 4646)	Novel Protein sim. GBank gi 2494162 sp Q10005 YRY1_CAEEL - HYPOTHETICAL 39.9 KD PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR	Contains protein domain (PF00226) - eph DnaJ domain	22278994, 22278995, 22278997, 60432049, 264259, 29331822, 33656970, 264509, 56182435, 264511, 265008, 60433356, 60433438, 55812038, 33109954, 21906754, 85658542, 87168474, 265011, 87168558, 265017, 265019, 264760, 264681, 18108351, 264369, 264288, 18108355, 264687, 264688, 21906765, 21906767, 21906768, 55811957, 35695917, 265021, 33657023, 18108362, 27468262, 55811576, 264631, 264555, 83373044, 87168518, 60432113, 22278002
2324	86633607 (4647, 4648)	Novel Protein sim. GBank gi 5419865 emb CAB46377.1  - (AL096732) hypothetical protein [Homo sapiens]	ATPase-associated	264592, 264593, 265020
2325	88165074 (4649, 4650)	Novel Protein sim. GBank gi 231885 sp P29881 CP4C_BLADI - CYTOCHROME P450 4C1 (CYPIVC1)	Contains protein domain (PF00067) - cyto450 Cytochrome P450	265006, 264759, 35695855, 56182323
2326	84390962 (4651, 4652)	Novel Protein sim. GBank gi 231885 sp P29881 CP4C_BLADI - CYTOCHROME P450 4C1 (CYPIVC1)	UNCLASSIFIED	
2327	88081648 (4653, 4654)	Novel Protein sim. GBank gi 4240227 dbj BAA74892.1  - (AB020676) KIAA0869 protein [Homo sapiens]	UNCLASSIFIED	264259, 264508, 264905, 264906, 264907, 264908, 265007, 264512, 264910, 264758, 265010, 264766, 264768, 264769, 33657023, 264693, 264628, 264631, 264634, 264638, 264639, 264488
2328	83388428 (4655, 4656)	Novel Protein sim. GBank gi 1245105 U46463  - glutamine repeat protein-1 [Mus musculus]	UNCLASSIFIED	60433438, 264595, 265017, 264766, 264692, 264629, 264635, 264636, 264638, 56182323, 60432113, 264566
2329	87604478 (4657, 4658)	Novel Protein sim. GBank gi 1169343 sp P42209 DIF6_MOUSE - DIF6 PROTEIN	UNCLASSIFIED	265017, 264685, 60432113, 264088
2330	87335396 (4659, 4660)	Novel Protein sim. GBank gi 5679136 gb AA046874.1 AF16093 - (AF160934) BcDNA.LD14189 [Drosophila melanogaster]	transport	265009
2331	86990463 (4661, 4662)	Novel Protein sim. GBank gi 2104452 emb CAB08779  - (Z95397) unknown [Schizosaccharomyces pombe]	ATPase-associated	35696286, 22278998, 29331824, 60424269, 265006, 265008, 265018, 264448, 264764, 21906765, 35695917, 35695855, 264636, 22279000, 264566
2332	87784182 (4663, 4664)	Novel Protein sim. GBank gi 2104452 emb CAB08779  - (Z95397) unknown [Schizosaccharomyces pombe]	UNCLASSIFIED	56182575, 56994075, 29331826, 29331828, 264107, 33657402, 87168559, 264683, 35695917, 265021, 33657023, 263976
2333	88206958 (4665, 4666)	Novel Protein sim. GBank gi 3879895 emb CAA92691.1  - (Z68318) cDNA EST CEIMSD62F comes from this gene; cDNA EST EMBL:CD7930 comes from this gene; cDNA EST EMBL:CD9493 comes from this gene; cDNA EST yk415e8.3 comes from this gene; cDNA EST yk415e8.5 comes from this gene; cDNA EST...	UNCLASSIFIED	
2334	94319788 (4667, 4668)	Novel Protein sim. GBank gi 4966270 gb AA052261.2  - (U97002) similar to acyl-CoA dehydrogenases and epoxide hydrolases; Pfam domain PF00441 (Acyl-CoA_dh), Score=57.4, E-value=1.7e-16, N=2; contains similarity to Pfam domain PF00702 (Hydrolase), Score=57.4, E- value=1e-13, N=1 [C...	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	56182575, 29331825, 21906768, 264636, 83373044

2335	80046103 (4669, 4670)	Novel Protein sim. GBank gll3283350 (AF062378) - calmodulin-binding protein SHA1 [Mus musculus]	Contains protein domain (PF00612) - IQ calmodulin-binding motif	struct	18108351, 21906769, 264555
2336	95196121 (4671, 4672)	Novel Protein sim. GBank gll1929056jemb[CAA72805] - (Y12090) putative 3,4-dihydroxy-2-butanone kinase [Lycopersicon esculentum]	kinase		264907, 35695917, 18108379
2337	95345810 (4673, 4674)	Novel Protein sim. GBank gll4495063jemb[CAB39181.1] - (Z85966) dJ108K11.3 (similar to yeast suppressor protein SRP40) [Homo sapiens]	UNCLASSIFIED		35696286, 22278999, 56182181, 29331825, 60424289, 56182435, 33657402, 55811386, 265017, 265018, 265019, 21906768, 35695917, 264691, 33657023, 33657109, 263972, 35696423, 35695855, 60432113
2338	87634045 (4675, 4676)	Novel Protein sim. GBank gll22246889jdbj[BAA20829] - (AB002372) KIAA0374 [Homo sapiens]	Contains protein domain (PF00323) - Mammalian defensin	UNCLASSIFIED	18108394, 29331822, 66714117, 60432289, 264906, 56182435, 265009, 60433438, 264596, 265010, 265019, 18108354, 264288, 264369, 55811957, 265021, 33657023, 263976, 55811576, 264632, 56182323, 264639
2339	85663319 (4677, 4678)	Novel Protein sim. GBank gll3873550jemb[CAA22127] - (AL033534) serine-rich protein [Schizosaccharomyces pombe]	UNCLASSIFIED		35696286, 264592, 264369, 264691, 264558
2340	90937716 (4678, 4680)				65274572, 22278994, 35696286, 22278997, 22278998, 264259, 29331822, 60432289, 29331828, 29331830, 265009, 33657402, 33109954, 265017, 265018, 264768, 264685, 21906769, 35695917, 264691, 264692, 35696423, 87188518, 22278900
2341	87775281 (4681, 4682)	Novel Protein sim. GBank gll3874563jemb[CAB02787] - (Z81042) similar to Yeast hypothetical protein YEY6 like; cDNA EST yk206h5.3 comes from this gene; cDNA EST yk206h5.5 comes from this gene; cDNA EST yk303h1.3 comes from this gene; cDNA EST yk303h1.5 comes from this gene; cDNA ...	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	264259, 264908, 264909, 264682, 22279000, 264488, 65274572, 22278995, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 66714117, 29331826, 60432289, 35696052, 264905, 264509, 29331830, 265006, 60170831, 60432229, 60433356, 87168474, 265017, 265018, 265019, 264448, 264369, 264288, 21906765, 21906766, 55811957, 35695917, 265020, 265022, 52644150, 33657023, 65274620, 33657109, 18108370, 18108376, 55810784, 35696423, 55811576, 264556, 264558, 18108385, 60432113, 264563, 264564, 264565, 264566, 264567, 264907, 264512, 265011, 264683
2343	87775448 (4685, 4686)	Novel Protein sim. GBank gll4929741jgbjAAD34131.1/AF151894 - (AF151894) CGI-136 protein [Homo sapiens]	UNCLASSIFIED		
2344	79953198 (4687, 4688)	Novel Protein sim. GBank gll2506307jispj13944[CA1C_CHICK - COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR (FIBROCHIMERIN)]	Contains protein domain (PF00092) - von Willebrand factor type A domain	UNCLASSIFIED	264758
2345	94319789 (4689, 4690)		collagen		264488, 264259, 66712502, 264759, 83373044, 264566

2346	94131820 (4691, 4692)	Novel Protein sim. GBank gl 1255411 (U53153) - one short region of weak similarity to S. cerevisiae protease A inhibitor 3 (SP:P01094) and another short region of weak similarity to S. cerevisiae glucose repression mediator protein (SP:P14922) [Caenorhabditis elegans]	Contains protein domain (PF00515) - TPR Domain		35696286, 22278998, 264259, 35696052, 29331828, 33657402, 60433356, 33109854, 87168559, 264603, 265019, 18108351, 264681, 264685, 21908768, 265021, 33657109, 55811576, 35695855, 264637, 52644332, 264557, 83373044, 22279000, 22279002
2347	85330367 (4693, 4694)				22278997, 264511, 264683, 264684, 264768, 264687, 264688, 264691, 264692, 55811576
2348	95186133 (4695, 4696)	Novel Protein sim. GBank gl 1929056[emb]CAA72805] - (Y12090) putative 3,4-dihydroxy-2-butanone kinase [Lycopodium obscurum]	kinase		18108394, 35696286, 264259, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264757, 264595, 264596, 264758, 265011, 264601, 264762, 18108351, 264764, 264288, 264766, 264768, 264689, 35695917, 264693, 264628, 18108370, 264629, 18108374, 35696423, 264631, 264635, 264636, 264637, 264638, 264639, 83373044, 18108385, 264567, 264486
2349	8776502 (4697, 4698)	Novel Protein sim. GBank gl 4884106[emb]CAB43254.1] - (AL050062) hypothetical protein [Homo sapiens]			35696052, 29146499, 264909, 264369
2350	88260594 (4699, 4700)				22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 33109954, 21906754, 265010, 87168559, 265018, 265019, 264761, 264681, 264288, 18108357, 21906766, 21906767, 264691, 264692, 35695855, 87168518, 22279000, 22279002, 264482
2351	86968042 (4701, 4702)	Novel Protein sim. GBank gl 728832[sp]P39189[ALU2_HUMAN - III] ALU SUBFAMILY SB WARNING ENTRY IIII	kinase		56182575, 264909, 265006, 264558
2352	87337186 (4703, 4704)	Novel Protein sim. GBank gl 731637[sp]P38760[YHH5_YEAST - HYPOTHETICAL 75.9 KD PROTEIN IN SPO13-ARG4 INTERGENIC REGION]	UNCLASSIFIED	264259, 264448	
2353	91638784 (4705, 4706)	Novel Protein sim. GBank gl 1346955[sp]P48809[RB27_DROME - HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN 27C (HNRNP 48) (HRP48.1)]	dna_ma_bind		29331826, 55812038, 265019, 264692, 264636
2354	87337199 (4707, 4708)	Novel Protein sim. GBank gl 731637[sp]P38760[YHH5_YEAST - HYPOTHETICAL 75.9 KD PROTEIN IN SPO13-ARG4 INTERGENIC REGION]	UNCLASSIFIED		29331824, 264908, 265006, 265008

2355	91638786 (4709, 4710)	Novel Protein sim. GBank gij4938503[embjCAB43861.1] - (AL078465) hnRNP-like protein [Arabidopsis thaliana]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	56994075, 22278996, 35696286, 22278999, 264259, 29331825, 29331826, 29331828, 29146498, 264905, 264908, 265008, 264758, 87168474, 265010, 265017, 264687, 21906765, 21906767, 21906769, 264691, 264692, 263987, 18108370, 87168518, 22279000
2356	95327688 (4711, 4712)	Novel Protein sim. GBank gij5138920[gbjAAD40377.1] - (AF092135) PTD014 [Homo sapiens]			52644507, 22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 60432289, 35696052, 29331828, 264908, 66712502, 264512, 265007, 265008, 60170831, 60432229, 60433356, 60433438, 264758, 52646317, 33109954, 21906754, 55811386, 87168474, 265017, 265018, 264605, 265019, 264681, 264682, 264448, 264369, 264288, 264686, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 60170615, 33657109, 27486264, 35695763, 55810764, 18108379, 35696423, 55811576, 35695855, 60170394, 56182323, 83373044, 18108385, 56526486, 264404, 60432113, 22279000, 264482, 264563, 264566, 264486, 264567
2357	87775459 (4713, 4714)	Novel Protein sim. GBank gij4929741[gbjAAD34131.1]AF15189 - (AF151894) CGI-136 protein [Homo sapiens]		UNCLASSIFIED	264488, 264769, 18108394, 264259, 29331822, 18108370, 18108374, 264510, 265017, 264482, 264563, 264762, 264565, 264566, 264369, 18108354
2358	87777078 (4715, 4716)	Novel Protein sim. GBank gij4218005 (AC006135) - putative vicilin storage protein (globulin-like) [Arabidopsis thaliana]		UNCLASSIFIED	22278997, 22278999, 264509, 264905, 264592, 18108351, 264681, 264682, 264769, 32833986, 18108374, 264556, 18108385, 264482
2359	87755859 (4717, 4718)	Novel Protein sim. GBank gij1086830 (U41264) - coded for by C. elegans cDNA yk208.5; coded for by C. elegans cDNA yk44g1.5; coded for by C. elegans cDNA yk12b7.5; coded for by C. elegans cDNA yk36g6.5; coded for by C. elegans cDNA yk208.5; coded for by C. elegans cDNA yk16g12.....		UNCLASSIFIED	35696286, 22278998, 264905, 264511, 265007, 265008, 60433438, 264288, 264686, 21906769, 265020, 264692, 35695855, 264558, 56526486, 264563
2360	80046125 (4719, 4720)	Novel Protein sim. GBank gij3881545[embjCAA93779] - (Z69804) cDNA EST yk428d5.3 comes from this gene; cDNA EST yk428d5.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	22278997, 29331826, 263981, 22279000
2361	94232181 (4721, 4722)	Novel Protein sim. GBank gij746487 (U23514) - No definition line found [Caenorhabditis elegans]			22278995, 22278999, 264512, 265009, 264757, 21906765, 65274620, 18108370, 60431528, 18108374, 264635, 60170394, 264482

2362	91721193 (4723, 4724)	Novel Protein sim. GBank gij1171093isp19706jmySB.ACACA - MYOSIN HEAVY CHAIN IB (MYOSIN HEAVY CHAIN IL)		UNCLASSIFIED	22278999, 264259, 29331822, 29331824, 60432289, 264509, 264512, 60432229, 60433356, 264448, 264682, 264683, 264369, 21906765, 21906768, 21906769, 60432113, 22279000, 22279002
2363	95006635 (4725, 4726)	Novel Protein sim. GBank gjl054065(emb CAA58337) - (X83413) U88 (Human herpesvirus 6)		UNCLASSIFIED	264907, 264629, 264635
2364	94827104 (4727, 4728)	Novel Protein sim. GBank gjl5639830jgb AAD45886.1 AF14601 - (AF146018) hydroxypyruvate reductase [Homo sapiens]	Contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases	reductase	264488, 18108394, 264887, 18108398, 22278996, 56994075, 35696286, 22278997, 22278998, 264259, 66714117, 29331825, 35696052, 264509, 264905, 264906, 264907, 264908, 66712502, 264909, 264511, 265006, 264512, 265007, 265008, 33657402, 264758, 21906754, 87168474, 265010, 87168559, 264603, 265017, 265018, 265019, 264760, 264762, 18108351, 264448, 264764, 264683, 264684, 264286, 18108355, 264766, 18108358, 264689, 18108359, 21906765, 21906766, 21906767, 35695917, 265020, 265021, 265022, 60170615, 52644150, 264691, 33657023, 264692, 18108364, 33657109, 18108368, 18108370, 18108374, 35696423, 35695855, 264635, 264556, 264557, 264639, 60170394, 83373044, 18108383, 18108384, 18108385, 18108388, 56528486, 264482, 264564, 264486
2365	94140746 (4729, 4730)	Novel Protein sim. GBank gjl1840045 (U49082) - transporter protein [Homo sapiens]		transport	22278998, 22278998, 22278999, 264907, 264909, 264910, 33657402, 264758, 264600, 264766, 264687, 264689, 21906765, 21906767, 21906768, 21906769, 265021, 33657023, 33657109, 83373044, 264566
2366	94312388 (4731, 4732)			UNCLASSIFIED	52644507, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 264907, 29331830, 264909, 264511, 265008, 33657402, 264595, 52646317, 265017, 265018, 265019, 264605, 264685, 264766, 264689, 21906766, 21906769, 35695917, 265020, 265021, 265022, 52644150, 35695855, 52644332, 18108385, 18108387, 264564, 264566
2367	94140910 (4733, 4734)	Novel Protein sim. GBank gjl1065457 (U40410) - C54G7.4 gene product [Caenorhabditis elegans]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	35696286, 21906768, 55810764, 65274791, 264567
2368	94322190 (4735, 4736)				264628



2369	94314334 (4737, 4738)	Novel Protein sim. GBank gjl5360901[dbj BAA82158.1] - saplens]	(AB029343) a-helix coiled-coil rod homologue [Homo sapiens]	struct	52644507, 52646842, 35696286, 264092, 264094, 52645080, 35698052, 264107, 29331830, 52644045, 265006, 265007, 265009, 52644296, 52644229, 264689, 21906765, 21906766, 35695917, 265020, 52644150, 263987, 33657109, 27486265, 35695763, 18108370, 263974, 18108374, 18108376, 52644332, 263981, 18108385
2370	79804120 (4739, 4740)			UNCLASSIFIED	264508, 264909, 264596
2371	57280406 (4741, 4742)			UNCLASSIFIED	264389
2372	87642413 (4743, 4744)			UNCLASSIFIED	263987, 263981
2373	87418611 (4745, 4746)	Novel Protein sim. GBank gjl4589582[dbj BAA76813.1] - saplens]	(AB023186) KIAA0989 protein [Homo sapiens]	UNCLASSIFIED	29331828, 265010, 265019, 35695917, 264634, 60432113
2374	94123665 (4747, 4748)	Novel Protein sim. GBank gjl5105131[dbj BAA80445.1] - saplens]	(AF000061) 246aa long hypothetical ribonuclease PH [Aeropyrum pernix]	UNCLASSIFIED	265006, 265007, 265008, 265009, 265011, 264766, 35695917, 35695855, 263981, 264557, 264555
2375	87731355 (4749, 4750)	Novel Protein sim. GBank gjl1351115[sp P47758 SRPB_MOUSE - SIGNAL RECOGNITION PARTICLE RECEPTOR BETA SUBUNIT (SR-BETA)]	Novel Protein sim. GBank gjl1351115[sp P47758 SRPB_MOUSE - SIGNAL RECOGNITION PARTICLE RECEPTOR BETA SUBUNIT (SR-BETA)]	UNCLASSIFIED	60432049, 29331824, 264907, 52644045, 264512, 60433358, 21908754, 52644296, 87168559, 264448, 21908765, 21908768, 21908769, 33657023, 18108368, 55811576, 52644332
2376	87613744 (4751, 4752)	Novel Protein sim. GBank gjl2645435 (AF007760) - CHD3 [Drosophila melanogaster]	Novel Protein sim. GBank gjl2645435 (AF007760) - CHD3 [Drosophila melanogaster]	ATPase associated PHD-finger	264259, 29331830, 264909, 264910, 265009, 60433438, 21908754, 265017, 265018, 265019, 264682, 264288, 264685, 21906767, 263972, 35695855, 87168518, 60432113
2377	95319689 (4753, 4754)	Novel Protein sim. GBank gjl5257005[gb AAD41239.1] - saplens]	Novel Protein sim. GBank gjl5257005[gb AAD41239.1] - saplens]	UNCLASSIFIED	264095, 29331822, 29147620, 29331824, 66714117, 29331825, 29331826, 29331828, 33656970, 29146498, 29146499, 264509, 265006, 265007, 265008, 265009, 60170931, 265010, 265011, 265018, 55811150, 18108351, 264764, 264288, 21908767, 21906768, 29148627, 29148629, 265021, 33657023, 33657109, 18108370, 18108374, 18108379, 35698423, 264556, 83373044, 18108385, 18108388, 55526486, 22279000, 22279002, 264563
2378	94137032 (4755, 4756)	Novel Protein sim. GBank gjl1072198 (U40942) - No definition line found [Caenorhabditis elegans]	Novel Protein sim. GBank gjl1072198 (U40942) - No definition line found [Caenorhabditis elegans]	UNCLASSIFIED	65274572, 56182575, 35696286, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 60432289, 29331827, 56182435, 264510, 265009, 60433356, 87168474, 265011, 265018, 264288, 21906765, 33657023, 264557, 56182323, 83373044, 18108385, 22278002, 264482
2379	65444324 (4757, 4758)	Novel Protein sim. GBank gjl3337357 (AC004481) - hypothetical protein [Arabidopsis thaliana]	Novel Protein sim. GBank gjl3337357 (AC004481) - hypothetical protein [Arabidopsis thaliana]	kinase	265017, 264288, 21906768

2380	86823082 (4758, 4760)	Novel Protein sim. GBank gi 4502839 ref NP_001845.1 pCOL1 - collagen, type XI, alpha 1	Contains protein domain (PF01410) - Fibrillar collagen C-terminal domain	collagen	264908, 264910, 265011
2381	87608241 (4761, 4762)	Novel Protein sim. GBank gi 4455609 emb CAB36555  - (AL031846) dJ742C19.5 (novel Chromobox protein) [Homo sapiens]	Contains protein domain (PF00385) - 'chromo' (CHRomain Organization Modifier) domain	helicase	56182575, 264091, 264093, 264259, 29331825, 264105, 264908, 60433356, 21906754, 265017, 265019, 264683, 264288, 264685, 264686, 264687, 264691, 264692, 264693, 55811576, 264636, 264567
2382	91225982 (4763, 4764)	Novel Protein sim. GBank gi 4325130 gb AAD17276  - (AF119716) dMI-2 protein [Drosophila melanogaster]	Contains protein domain (PF00628) - PHD-finger	transport	29331824, 60432289, 264905, 264596, 21906754, 264789, 265022, 264693, 263967, 33657109, 264629, 264631, 264558, 83373044, 60432113, 264482
2383	87442841 (4765, 4766)	Novel Protein sim. GBank gi 1902982 dbj BAA19005  - (D89049) lectin-like oxidized LDL receptor [Bos taurus]	Contains protein domain (PF00059) - Lectin C-type domain	glycoprotein	265009, 21906765, 21906766
2384	95354766 (4767, 4768)	Novel Protein sim. GBank gi 2482851 (AF016252) - Spinophilin [Rattus norvegicus]	Contains protein domain (PF00359) - PDZ domain (Also known as DHR or GLGF).	struc	264488, 52644507, 52645156, 52646365, 35696286, 22278999, 52645080, 29331824, 29331826, 35696052, 29331828, 264908, 264828, 52644045, 265006, 265008, 265009, 33109954, 33657084, 52644296, 265011, 265017, 265018, 264683, 52644229, 21906765, 21906767, 21906768, 265020, 52644150, 33657023, 264693, 65274620, 52645129, 33657109, 33657182, 27486261, 27486262, 27486264, 33657349, 27486265, 35695763, 18108374, 35695855, 264634, 264555, 264556, 264557, 52644332, 264558, 264559, 18108385, 22279000, 22279002
2385	95419485 (4769, 4770)		UNCLASSIFIED		264488, 52644507, 52645156, 264887, 52646365, 22278995, 22278996, 22278997, 22278998, 264259, 52645080, 29331822, 29331826, 35696052, 52644045, 265006, 265007, 265008, 265009, 264910, 60432229, 60433356, 52646317, 21908754, 265019, 264448, 264683, 264686, 264687, 264689, 21906765, 21906766, 21906767, 21906769, 55811957, 265021, 265022, 264690, 264691, 264692, 65274620, 33657109, 18108370, 264631, 52644332, 22279000, 22278002, 264563, 264565, 264567
2386	94742649 (4771, 4772)	Novel Protein sim. GBank gi 4929699 gb AAD34110.1 AF151873 CGI-115 protein [Homo sapiens]		glycoprotein	264488, 22278995, 22278996, 22278997, 264259, 29146498, 264112, 264511, 60170831, 60432229, 264595, 60433438, 87168474, 87168559, 264682, 21906765, 21906766, 21906767, 21906769, 29148629, 35695917, 265021, 264690, 33657109, 264628, 18108376, 83373044, 60432113, 22279000, 264564, 264566, 264487

2387	14997990 (4773, 4774)			UNCLASSIFIED	264634
2388	11424604 (4775, 4776)			UNCLASSIFIED	264595
2389	95310650 (4777, 4778)	Novel Protein sim. GBank gi 4758058 ref NP_004372.1 pCREB - cAMP responsive element binding protein-like 1	Contains protein domain (PF00170) - bZIP transcription factor		264488, 22278998, 22278999, 264509, 264905, 264906, 264907, 264908, 264909, 265006, 264511, 264512, 264910, 264591, 21908754, 264601, 264604, 264761, 18108351, 264764, 264288, 264766, 264768, 264769, 21906765, 21906768, 264692, 264693, 35696423, 264635, 264636, 264555, 83373044, 22279000, 264486
2390	94320912 (4778, 4780)	Novel Protein sim. GBank gi 1644239 dbj BAA12223  - (D84103) mitochondrial DNA polymerase gamma [Homo sapiens]	Contains protein domain (PF00476) - DNA polymerase family A	polymerase	52644507, 56182575, 22278995, 35696286, 22278996, 22278997, 22278999, 29331822, 29331825, 29331826, 35696052, 264905, 52644045, 265009, 264758, 264759, 33109954, 52644296, 85658542, 265011, 265017, 265018, 264605, 52644229, 21906765, 21906767, 21906768, 21906769, 35695917, 52644150, 33657023, 33657109, 33657349, 35695763, 18108370, 18108374, 18108376, 35696423, 35695855, 264555, 52644332, 56182323, 60170394, 83373044, 56526486
2391	80036194 (4781, 4782)			UNCLASSIFIED	263976
2392	94245016 (4783, 4784)	Novel Protein sim. GBank gi 4240169 dbj BAA74863.1  - (AB020647) KIAA0840 protein [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat	nuclease	35696286, 35696052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264910, 265009, 264591, 264758, 264600, 264604, 264762, 264448, 264764, 264369, 264766, 264768, 264769, 264689, 35695917, 264629, 18108374, 263978, 35696423, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 60170394, 264639, 264565, 264486

2393	95302633 (4785, 4786)	Novel Protein sim. GBank gi14506667 ref NP_000993.1 pRPLP - ribosomal protein, large, P0	Contains protein domain (PF00466) - Ribosomal protein L10	- ribosomalprot	18108392, 60424179, 264489, 18108394, 18108397, 22278995, 56994075, 35696286, 22278996, 22278997, 22278999, 264093, 60432049, 264259, 29331822, 29147620, 20281099, 29331824, 29331825, 68714117, 60432289, 29331826, 29331827, 29331828, 35696052, 29148499, 264508, 264509, 264905, 264907, 284908, 66712502, 52644045, 264828, 284909, 56182435, 264112, 264113, 264510, 265006, 264511, 265007, 265008, 265009, 264910, 264591, 264593, 60433356, 264595, 60433438, 52646317, 33109954, 21906754, 55811386, 265010, 265011, 265017, 265018, 265019, 264681, 264762, 18108351, 264763, 264682, 264764, 264683, 264369, 264288, 18108354, 264768, 264686, 264687, 264688, 264689, 18108359, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 29148629, 29148784, 35695917, 265021, 265022, 33657023, 264692, 264693, 18108364, 33657109, 18108368, 27486261, 27486262, 33657349, 35695763, 18108370, 263972, 264629, 18108374, 263977, 18108376, 263978, 55810764, 35696423, 35695855, 264634, 60431850, 264555, 264637, 264557, 263981, 264558, 18108381, 60170394, 35696286, 22278997, 22278998, 56182181, 35696052, 265006, 264592, 55811388, 265010, 265011, 265017, 265019, 264448, 264683, 264288, 21906765, 21906768, 21906769, 55811957, 35695917, 33657023, 65274620, 33657182, 33657349, 35695763, 18108374, 18108376, 55810764, 55811576, 35696423, 60170394, 18108385, 264564, 264566, 264567
2394	94323266 (4787, 4788)	Novel Protein sim. GBank gi14159888 (AC004908) - zinc finger protein from gene of uncertain exon structure; similar to Q99676 (PID:g3025333) [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	- dna_ma_bind	
2395	95287212 (4789, 4790)	Novel Protein sim. GBank gi15712756 gb AAD47636.1 AF16079 - (AF16079) calcium transporter CaT1 [Rattus norvegicus]		- dna_ma_bind	264259, 29331824, 264910, 264288, 265021, 83373044, 18108387, 264563, 264566

2386	85086700 (4791, 4792)	Novel Protein sim. GBank gi 106322 pir B3-087 - hypothetical protein (L1H 3' region) - human	Contains protein domain (PF00560) - Leucine Rich Repeat	nuclease	52646365, 18108397, 56182575, 35696286, 22278997, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264106, 264907, 29331830, 66712502, 264110, 60170831, 264591, 33657402, 60433438, 55812038, 33109954, 21906754, 33657084, 87168474, 265017, 265018, 265019, 264760, 264448, 264288, 264786, 52644229, 21906766, 21906767, 265020, 265021, 60170615, 264692, 33657023, 65274620, 52645129, 33657182, 27486262, 27486264, 27486265, 264629, 18108374, 35698423, 35698555, 264631, 264556, 52644332, 264558, 83373044, 18108388, 87168518, 22279002, 264482
2397	87280854 (4793, 4794)				52644507, 52645156, 56182575, 264259, 29147620, 264805, 264907, 264908, 264809, 264910, 264758, 52644296, 264603, 264604, 264762, 264681, 264764, 18108357, 264769, 21906768, 264693, 264628, 264635, 264638, 264639, 264584
2398	88047689 (4795, 4796)	Novel Protein sim. GBank gi 3258609 (AC005178) - H53_GS1 [Homo sapiens]		UNCLASSIFIED	
2399	87738965 (4797, 4798)	Novel Protein sim. GBank gi 786117 (L41834) - nuclear protein [Ensis minor]		UNCLASSIFIED	56994075, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 66714117, 60432289, 264806, 29331830, 56182435, 264112, 264910, 33109954, 21906754, 87168474, 264600, 265017, 265018, 265019, 264764, 264765, 21906765, 21906766, 21906767, 21906769, 35695917, 265020, 265022, 60170615, 33657023, 18108370, 18108374, 264556, 60170394, 264558, 87168518, 22279000, 22279002, 264564, 264566, 264487
2400	91214116 (4799, 4800)	Novel Protein sim. GBank gi 2352822 gb AAB69285.1  - (AF008945) glucose-6-phosphatase [Haplochromis nubilus]		phosphatase	21806766, 52646842, 56994075, 33657182, 27486262, 52644296, 265017

2401	91214118 (4801, 4802)	Novel Protein sim. GBank gij2352822[gblAA689285.1] - (AF008945) glucose-6-phosphatase [Haplochromis nubilus]		phosphatase	52644507, 52645156, 52644228, 264688, 21906764, 21906765, 52646365, 52646842, 21906766, 21906767, 21906768, 22278995, 35695917, 56994075, 35696286, 22278996, 22278997, 265020, 22278998, 22278999, 264259, 33657023, 52645080, 264693, 29331824, 33657109, 52645129, 29331826, 33657182, 29331827, 35696052, 27486261, 27486262, 33656970, 33657349, 27486265, 35695763, 264106, 264905, 35696423, 35695855, 265006, 265007, 265008, 265009, 264637, 52644332, 55812038, 52646317, 18108385, 52644296, 87168474, 265010, 87168558, 60432113, 265017, 265018, 265019, 264563, 264288
2402	91221408 (4803, 4804)	Novel Protein sim. GBank gij4680258[gblAAD27832.1]AF12185 - (AF121859) sorting nexin 8 [Homo sapiens]			264907, 264908, 264909, 264566
2403	94135432 (4805, 4806)	Novel Protein sim. GBank gij4929575[gblAAD34048.1]AF15181 - (AF151811) CGI-53 protein [Homo sapiens]	Contains protein domain (PF00062) - C-type lysozyme/alpha-lactalbumin family		22278999, 35696052, 265018, 264686, 264693, 83373044, 264567
2404	95312605 (4807, 4808)	Novel Protein sim. GBank gij2315786 (AF016685) - similar to short chain-type dehydrogenases [Caenorhabditis elegans]	Contains protein domain (PF00106) - short chain dehydrogenase	dehydrogenase	35696286, 29331826, 35696052, 265008, 265018, 21906769, 264564
2405	94311851 (4808, 4810)	Novel Protein sim. GBank gij464178[dbj BAA03581] - (D14853) polyprotein [Hepatitis C virus]		UNCLASSIFIED	35696286, 29331822, 265007, 21906754, 265017, 265018, 265019, 264763, 264369, 21906765, 35695917, 265020, 265021, 52644150, 264693, 35695855, 264632, 52644332, 22279002
2406	88094501 (4811, 4812)	Novel Protein sim. GBank gij2773363 (AF041382) - microtubule binding protein D-CLIP-190 [Drosophila melanogaster]	Contains protein domain (PF01302) - CAP-Gly domain	struct	52646842, 22278994, 22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264093, 60432049, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 264102, 264106, 264906, 52644045, 265007, 265008, 265009, 264910, 264592, 60433356, 60433438, 33109954, 265010, 265011, 265018, 265019, 264369, 264685, 264686, 21906768, 21906769, 52644150, 264693, 52645129, 264628, 35696423, 264632, 56182323, 264639, 22279000, 22279002, 264563
2407	79465005 (4813, 4814)			UNCLASSIFIED	264685, 264686
2408	87391503 (4815, 4816)	Novel Protein sim. GBank gij423442[pirl S33513 - gene Fil protein - mouse		UNCLASSIFIED	264910, 265010, 264448, 264557

2409	84741770 (4817, 4818)	Novel Protein sim. GBank gi 1176801 sp P45968 YNZ6_CAEEL - HYPOTHETICAL 20.8 KD PROTEIN T09A5.6 IN CHROMOSOME III		UNCLASSIFIED	22278995, 22278996, 22278997, 264097, 29331822, 29331824, 29331827, 29146498, 52644045, 60433438, 33657084, 87168474, 264760, 21906767, 29148627, 29148629, 52644150, 33657023, 263967, 20281069, 18108374, 20281071, 56182323, 83373044, 18108385, 87168518 22278998, 264259
2410	87604660 (4819, 4820)	Novel Protein sim. GBank gi 4966262 gb AAC48052.2  - (UB4849) Contains similarity to Plam domain: PF00646 (F- box), Score=28.7, E-value=4.3e-05, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00080) - Copper/zinc superoxide dismutase (SODC)		
2411	87534633 (4821, 4822)	Novel Protein sim. GBank gi 3114713 (AF061346) - Edp1 protein [Mus musculus]		inf	29331824, 29331827, 29331828, 264764, 264369, 33657109, 56182323
2412	87778332 (4823, 4824)	Novel Protein sim. GBank gi 5410336 gb AAD43038.1  - (AF106685) myelin gene expression factor 2 [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	22278998, 29331827, 264907, 265011, 265017, 265018, 265019, 18108351, 21906766, 265020, 33657109, 264559, 18108385
2413	94133820 (4825, 4826)	Novel Protein sim. GBank gi 5262705 emb CAB45778.1  - (AL080214) hypothetical protein [Homo sapiens]	Contains protein domain (PF00038) - Intermediate filament proteins	struct	264488, 264259, 29331826, 264508, 264905, 264509, 264906, 264907, 264908, 264510, 264511, 264512, 265008, 265009, 264910, 265011, 264682, 264764, 264766, 264686, 264768, 264689, 265021, 33657023, 18108370, 264628, 35695855, 264632, 264634, 264635, 264636, 83373044, 264563, 264564, 264565, 264566, 264567, 264486
2414	94312580 (4827, 4828)	Novel Protein sim. GBank gi 1082340 pir J52863 - DNA- binding protein R kappa B - human		ubiquitin	52645156, 52646842, 52646385, 18108398, 56182575, 22278994, 22278995, 56994075, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 33656970, 29331830, 264909, 60433356, 33657402, 264594, 52646317, 21906754, 33657084, 265010, 87168559, 265017, 265018, 265019, 264369, 264684, 264687, 264688, 56181562, 21906764, 264689, 21906765, 21906766, 21906767, 29148627, 21906769, 265020, 265021, 60170615, 33657023, 264693, 52645129, 33657109, 33657182, 27486261, 27486262, 27486265, 33657349, 18108370, 60431528, 264629, 18108374, 18108378, 55810764, 264636, 52644332, 264638, 264558, 56182323, 83373044, 18108385, 87168518, 22279002

2415	8808002 (4829, 4830)	Novel Protein sim. GBank gi 423915 pir A45439 - myosin I heavy chain - rat	Contains protein domain (PF00063) - Myosin head (motor domain)	struct	264259, 264908, 60433356, 33657402, 21906754, 265018, 264687, 264689, 21906768, 55811957, 265021, 264690, 264691, 33657023, 264693, 35696423, 56182323, 56526486
2416	94118356 (4831, 4832)	Novel Protein sim. GBank gi 3025445 (AC004528) - R32184_1 [Homo sapiens]			264638
2417	87733334 (4833, 4834)	Novel Protein sim. GBank gi 1084944 pir S54495 - hypothetical protein YPR021c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00153) - Mitochondrial carrier proteins		264094, 29331822, 29331824, 29331827, 264369
2418	94234349 (4835, 4836)	Novel Protein sim. GBank gi 1176572 sp P45895 YNA4_CAEEL - HYPOTHETICAL 91.0 KD PROTEIN PAR.2.4 IN CHROMOSOME III	Contains protein domain (PF00411) - UNCLASSIFIED Ribosomal protein S11		56994075, 264091, 264259, 29331824, 29331825, 60432289, 29331828, 264905, 264907, 264511, 265009, 60432229, 21906754, 87168559, 265019, 264682, 21906768, 21906769, 265020, 265021, 33657023, 65274620, 18108370, 55811576, 264634, 60170394, 18108385, 22279000, 22279002, 264566
2419	82374249 (4837, 4838)	Novel Protein sim. GBank gi 284006 pir S18732 - autoantigen, 64K - human		struct	264569, 264762, 264448, 264691, 264631, 264634, 264555, 264558, 264638, 264658
2420	94844244 (4839, 4840)	Novel Protein sim. GBank gi 1076211 pir S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii		UNCLASSIFIED	29331824, 29331825, 29331828, 60432229, 33109954, 85658542, 87168474, 265018, 264288, 265020, 264584
2421	87805345 (4841, 4842)			UNCLASSIFIED	264909, 264768, 264638
2422	88084714 (4843, 4844)	Novel Protein sim. GBank gi 2224567 dbj BAA207721 - (AB002311) KIAA0313 [Homo sapiens]	Contains protein domain (PF00617) - RasGEF domain	transport	18108392, 18108394, 18108398, 264906, 265006, 265010, 18108351, 18108374, 18108385
2423	88058390 (4845, 4846)	Novel Protein sim. GBank gi 4505153 ref NP_002392.1 pMEKK - MAP/ERK kinase kinase 3	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264259, 60432049, 29331822, 29331826, 60432289, 29331828, 265008, 265009, 60433356, 21906754, 265017, 265018, 265019, 21906766, 21906768, 21906769, 265020, 265021, 20281149, 263971, 60432113
2424	94854047 (4847, 4848)	Novel Protein sim. GBank gi 2988398 (AC004381) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	56182575, 35696286, 22278997, 60432049, 264259, 29331826, 29331828, 264905, 66712502, 29331830, 60433356, 265011, 265019, 264766, 21906768, 55811957, 264692, 33657023, 33657109, 55811576, 56182323, 83373044, 18108385, 18108388, 60432113, 22279000
2425	87415981 (4849, 4850)	Novel Protein sim. GBank gi 2077932 dbj BAA19879 - (D86558) Protein Kinase [Rattus norvegicus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264634
2426	87613945 (4851, 4852)	Novel Protein sim. GBank gi 2039368 gb AAB53003.1 - (U94619) circulating cathodic antigen [Schistosoma mansoni]		UNCLASSIFIED	22278996, 22278998, 264259, 264102, 264512, 265008, 21906767, 18108370, 18108374, 263976



2427	87622693 (4853, 4854)	Novel Protein sim. GBank gi 4680695 gb AAD27737.1 AF13296 - (AF13296) CGI-28 protein [Homo sapiens]	Contains protein domain (PF00573) - Ribosomal protein L4/L1 family	ribosomalprot	264259, 20281099, 35696052, 265008, 264594, 265011, 264760, 18108351, 264682, 264683, 264369, 264684, 264686, 264687, 264689, 21906766, 264691, 264692, 18108374, 18108377, 264557, 264639, 18108385
2428	85732889 (4855, 4856)	Novel Protein sim. GBank gi 1537070 (U63840) - nucleoprotein p54 [Rattus norvegicus]			22278996, 22278999, 35696052, 21906754, 264288, 21906765, 21906768, 21906769, 35695917, 265020, 263972, 22279002
2429	87769276 (4857, 4858)	Novel Protein sim. GBank gi 601931 (M94316) - neurofilament-H [Oryctolagus cuniculus]	Contains protein domain (PF00711) - Beta defensins	UNCLASSIFIED	22278999, 29331824, 264806, 264909, 264511, 265009, 21906754, 265017, 265018, 265019, 264448, 264683, 264288, 21906765, 21906768, 265021, 264693, 18108381
2430	86948827 (4859, 4860)	Novel Protein sim. GBank gi 3860729 emb CAA14630  - (AJ235270) CELL DIVISION PROTEIN FTSJ (fts.j) [Rickettsia prowazekii]	Contains protein domain (PF01728) - FtsJ cell division protein	UNCLASSIFIED	264112, 264691
2431	87649884 (4861, 4862)	Novel Protein sim. GBank gi 3876367 emb CAA93287  - (Z69360) Weak similarity to Eimeria thrombospondin (PIR Acc. No. A45517); cDNA EST EMBL:M9266 comes from this gene; cDNA EST yk295b9.5 comes from this gene [Caenorhabditis elegans]		protease	22278996, 29331827, 35696052, 29146499, 264905, 264906, 264681, 264288, 264689, 21906765, 264692, 35696423 264634, 264558
2432	80083033 (4863, 4864)	Novel Protein sim. GBank gi 2224593 dbj BAA20784  - (AB002324) KIAA0326 [Homo sapiens]	Contains protein domain (PF000036) - Zinc finger, C2H2 type	transcriptfactor	264569, 264905, 265018, 264762, 264683, 264691, 264556, 264557, 264639, 264558 264563
2433	80055092 (4865, 4866)	Novel Protein sim. GBank gi 1263289 (U47856) - fibroin-4 [Araneus diadematus]		UNCLASSIFIED	264555
2434	19520148 (4867, 4868)	Novel Protein sim. GBank gi 3641352 (AF091234) - putative transcription factor [Mus musculus]		UNCLASSIFIED	265008, 264758, 265010, 264689, 27486261, 263972, 18108374, 18108381
2435	20759044 (4869, 4870)	Novel Protein sim. GBank gi 3860014 (AF091088) - unknown [Homo sapiens]	Contains protein domain (PF01256) - Uncharacterized protein family UPF0031	UNCLASSIFIED	29331828, 265007, 265009, 265017, 264760, 264685, 264693, 264565
2436	88044008 (4871, 4872)	Novel Protein sim. GBank gi 3860014 (AF091088) - unknown [Homo sapiens]		UNCLASSIFIED	29331826, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264511, 265006, 264512, 264910, 265009, 264591, 33657402, 21906754, 265011, 264760, 264764, 264685, 264686, 264768, 35695917, 33657023, 264693, 264631, 264632, 56182323, 264558, 83373044, 264563, 264564, 264565, 264566, 264567

2439	94850650 (4877, 4878)	Novel Protein sim. GBank gi 4263519 gb AA015345  - (AC004044) small nuclear riboprotein Sm-D1 [Arabidopsis thaliana]	Contains protein domain (PF01423) - Sm protein	UNCLASSIFIED	60424179, 18108397, 56182575, 56181886, 56994075, 22278996, 35696286, 22278997, 22278999, 264259, 52645080, 29331822, 56182181, 29331824, 60424269, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 29146498, 264509, 264905, 264906, 52644045, 60431735, 33109954, 21905754, 33657084, 55811386, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264448, 264288, 264768, 52644229, 58181562, 21906764, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 33657023, 33657109, 33657182, 27486262, 27486264, 33657349, 27486265, 35695763, 18108370, 60431528, 263977, 55810764, 35696423, 65274791, 35695855, 60431850, 56182323, 60432113, 22278000, 22279002, 264567
2440	87641733 (4879, 4880)			UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 35696052, 264508, 264907, 264510, 265018, 265019, 264448, 264369, 265020, 265021, 56182323, 264639, 22278002
2441	87623914 (4881, 4882)	Novel Protein sim. GBank gi 3024889 sp P56524 Y288_HUMAN - HYPOTHETICAL PROTEIN KIAA0288 (HA6116)		UNCLASSIFIED	264488, 264629, 18108374, 264564
2442	87273590 (4883, 4884)	Novel Protein sim. GBank gi 4506013 ref NP_002703.1 ppPPP1 - protein phosphatase 1, regulatory subunit 7	Contains protein domain (PF00560) - Leucine Rich Repeat	UNCLASSIFIED	22278996, 22278999, 29331822, 264768, 264693
2443	84305849 (4885, 4886)	Novel Protein sim. GBank gi 1170658 sp Q02975 KID1_RAT - RENAL TRANSCRIPTION FACTOR KID-1 (TRANSCRIPTION FACTOR 17)	Contains protein domain (PF01352) - KRAB box	transcriptfactor	264906
2444	88086345 (4887, 4888)	Novel Protein sim. GBank gi 4758824 ref NP_004280.1 pNRF3 - nuclear factor (erythroid-derived 2)-like 3	Contains protein domain (PF00170) - bZIP transcription factor	transcriptfactor	264259, 18108382, 18108383, 18108385, 22278000
2445	87338636 (4889, 4890)	Novel Protein sim. GBank gi 2135950 pir I S58222 - PQ-rich protein - human			264259, 35696052, 264369, 18108361
2446	88059293 (4891, 4892)	Novel Protein sim. GBank gi 4753887 emb CAA05409.2  - (AJ002424) p65 protein [Rattus norvegicus]	Contains protein domain (PF00095) - WAP-type (Whey Acidic Protein) 'four-disulfide core'	proteaseinhib	265011, 264689, 33657023, 263981, 18108385
2447	94845149 (4893, 4894)	Novel Protein sim. GBank gi 4885813 ref NP_005409.1 pST5j - suppression of tumorigenicity 5		cadherin	56182575, 264259, 29331824, 29331825, 29331827, 60433356, 60433438, 264758, 265018, 264692, 65274620, 60431528, 65274791, 56182323

2448	87749680 (4895, 4896)			UNCLASSIFIED	22278996, 22278997, 22278998, 28331826, 35696052, 264107, 264110, 87168474, 87168559, 18108351, 21906767, 21906769, 27486262, 263976
2449	87869075 (4897, 4898)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		cadherin	264259, 264828, 265007, 264595, 265021, 56526486
2450	86597784 (4899, 4900)	Novel Protein sim. GBank		UNCLASSIFIED	264906
2451	91014563 (4901, 4902)	gi 1710021 sp P35290 RB24_MOUSE - RAS-RELATED PROTEIN RAB-24 (RAB-16)	Contains protein domain (PF00071) - Ras family	glycoprotein	264003, 29331822, 29331824, 29331825, 66714117, 29331826, 29331828, 35696052, 264907, 66712502, 29331830, 264910, 265009, 264758, 265017, 265018, 264762, 264448, 264288, 21906767, 265021, 33657023, 264693, 33657109, 263969, 83373044, 18108385
2452	91230509 (4903, 4904)	Novel Protein sim. GBank gi 1504034 pb BAA13216  - (D86980) KIAA0227 [Homo sapiens]		isomerase	264102, 264112, 264688, 263972, 18108374, 83373044, 264563
2453	84201088 (4905, 4906)	Novel Protein sim. GBank gi 2880079 (AC004142) - similar to murine leucine-rich repeat protein; possible role in neural development by protein-protein interactions; 93% similarity to D48802 (PID:g1369908) [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat	ngrecep	264509, 264512, 18108385
2454	95310691 (4907, 4908)	Novel Protein sim. GBank gi 1076802 pir S49915 - extensin like protein - maize	Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	263994, 66714117, 29331827, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 265009, 264910, 264591, 264758, 264759, 265010, 265011, 264603, 264604, 264760, 264761, 264762, 18108351, 264764, 264765, 264766, 264686, 264768, 264769, 264534, 264691, 264692, 33657023, 264693, 33657109, 264628, 263978, 35695855, 264634, 264635, 264637, 264638, 264639, 83373044, 18108385, 264563, 264564, 264486
2455	95280301 (4909, 4910)	Novel Protein sim. GBank gi 543817 sp P35585 AP47_MOUSE - CLATHRIN COAT ASSEMBLY PROTEIN AP47 (CLATHRIN COAT ASSOCIATED PROTEIN AP47) (GOLGI ADAPTOR AP-1 47 KD PROTEIN) (HA1 47 KD SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN ASSEMBLY PROTEIN COMPLEX 1 MEDIUM CHAIN)	Contains protein domain (PF00928) - Adaptor complexes medium subunit family	glycoprotein	264488, 22278996, 264259, 35696052, 264905, 264906, 264907, 264908, 264909, 264510, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264758, 265019, 264760, 264681, 18108351, 264683, 264764, 18108354, 264766, 264768, 264769, 264689, 21906766, 21906767, 21906769, 29148628, 35695917, 265020, 265022, 33657023, 33657109, 18108370, 264628, 264629, 264631, 264632, 264635, 56182323, 60170394, 18108385, 264563, 264564, 264566, 264567
2456	88166700 (4911, 4912)	Novel Protein sim. GBank gi 2588630 (AC003079) - Ankyrin- like; 54% similar to 2022340A (NID:g1092123) in exons spanning 43974 to 11551 of clone. [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	kinase	264693

2457	94118375 (4913, 4914)	Novel Protein sim. GBank gij3025447 (AC004528) - R32184_3 [Homo sapiens]		UNCLASSIFIED	55181686, 264905, 264907, 264511, 264596, 55811386, 264682, 264684, 264685, 264687, 264691, 33857023, 264693, 35695855, 264636, 264555, 56182323, 264558, 56526488, 264563
2458	85675304 (4915, 4916)	Novel Protein sim. GBank gij2384942 (AF022985) - Similar to collagen [Caenorhabditis elegans]		UNCLASSIFIED	264691, 264693, 264634, 264559
2459	87551913 (4917, 4918)	Novel Protein sim. GBank gij5441942 [gbjAAD3187.1] AC004997 supported by mouse EST AA538043 (NID: g2284036) [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	transport	27486265
2460	94315289 (4919, 4920)	Novel Protein sim. GBank gij4929701 [gbjAAD34111.1] AF15187 - (AF151874) CGI-116 protein [Homo sapiens]		kinase	65274572, 35696286, 22278996, 22278997, 60432049, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 35656970, 29146499, 264102, 264109, 60433438, 265017, 265018, 265019, 264288, 21906765, 21906766, 21906769, 35695917, 265020, 264691, 33657023, 27486261, 18108374, 35695855, 87168518, 60432113
2461	87645147 (4921, 4922)	Novel Protein sim. GBank gij4426962 [gbjAAD20633] - (AF126062) Arf-like 2 binding protein BART1 [Homo sapiens]		UNCLASSIFIED	264259, 29331828, 264910, 18108351, 18108370, 18108374
2462	86990002 (4923, 4924)	Novel Protein sim. GBank gij5420387 [embjCAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]			264909, 264758, 264684, 18108374, 264637, 18108385
2463	84388543 (4925, 4926)	Novel Protein sim. GBank gij5052516 [gbjAAD38588.1] AF14561 - (AF145613) BcDNA.GH03108 [Drosophila melanogaster]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	264681, 264566
2464	91219957 (4927, 4928)	Novel Protein sim. GBank gij5410300 [gbjAAD43021.1] - (AF100757) COP9 complex subunit 4 [Homo sapiens]	Contains protein domain (PF01399) - PCI domain	protease	264489, 52646842, 22278995, 35696286, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 29331827, 35696052, 29331828, 264509, 52644045, 264510, 264511, 264512, 265008, 60170831, 264593, 52646317, 33109954, 33657084, 265017, 265018, 265019, 264762, 264448, 264764, 264288, 264766, 21906765, 21906768, 21906769, 21906769, 265021, 33657023, 33657109, 18108370, 18108381, 60170394, 18108385, 22279002, 264486

2465	85357483 (4929, 4930)	Novel Protein sim. GBank gi 4506401 ref NP_002871.1 pRAF1 - v-raf-1 murine leukemia viral oncogene homolog 1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	oncogene	18108392, 52644507, 52645156, 52646365, 22278994, 22278995, 35698286, 22278996, 22278998, 264259, 29331822, 29331824, 29331825, 60424269, 60432289, 29331827, 35696052, 29331828, 264907, 29331830, 52644045, 264909, 56182435, 264511, 265007, 265008, 265009, 264910, 33657402, 60433438, 55812038, 21908754, 33109954, 265010, 265011, 87168559, 264600, 265017, 265018, 265019, 18108351, 264369, 264288, 264685, 264767, 21908765, 21908787, 21908788, 55811957, 35695917, 265020, 265021, 60170615, 52644150, 33857023, 52645129, 33657109, 27486261, 27486264, 35695763, 264628, 263972, 18108374, 35695855, 264636, 264637, 60170394, 56526486, 87168518, 60432113, 264563, 264564, 264566, 264487 264369
2466	85681388 (4931, 4932)	Novel Protein sim. GBank gi 4321619 gb AAD15788.1  - (AF051098) seven transmembrane domain orphan receptor [Mus musculus]			
2467	88059465 (4933, 4934)	Novel Protein sim. GBank gi 3513300 (AC005595) - F16601.1, partial CDS [Homo sapiens]		UNCLASSIFIED	56994075, 264908, 21908788, 33857023
2468	87614696 (4935, 4936)	Novel Protein sim. GBank gi 2143455 pir j158106 - gene DMR-N9 protein - mouse (fragment)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	29331824, 52644045, 265008, 264910, 265019, 21908785, 21908789, 265021 264288, 264628
2469	86294397 (4937, 4938)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264592, 264692, 264555, 264556, 264557, 264558, 264559, 18108385, 264482
2470	80223831 (4939, 4940)	Novel Protein sim. GBank gi 5419882 emb CAB46424.1  - (AL086749) DKFZp434G153 [Homo sapiens]		UNCLASSIFIED	65274572, 35896286, 29331827, 265007, 264592, 33109954, 265018, 265019, 264288, 55811957, 265020, 264693, 55811576, 56182323
2472	95060811 (4943, 4944)	Novel Protein sim. GBank gi 4929747 gb AAD34134.1 AF15189 - (AF151897) CGI-139 protein [Homo sapiens]		UNCLASSIFIED	264092, 60432049, 29331825, 60433356, 265010, 265011, 18108351, 264764, 264288, 264692, 65274620, 18108370, 18108372, 18108374, 264634, 18108385
2473	95421509 (4945, 4946)	Novel Protein sim. GBank gi 4539009 emb CAB39630.1  - (AL049481) putative protein [Arabidopsis thaliana]			60424179, 65274572, 22278999, 60424269, 29331826, 265008, 60433356, 60433438, 265010, 18108351, 264448, 264288, 264687, 264689, 265021, 264692, 65274620, 60431528, 65274791, 264556, 56182323, 60432113
2474	94315616 (4947, 4948)	Novel Protein sim. GBank gi 3252827 (AC004382) - Unknown gene product [Homo sapiens]			65274572, 56994075, 264259, 29331826, 60170831, 265017, 265018, 265019, 264683, 264369, 265020, 264693, 264563, 264564

2475	94321683 (4949, 4950)	Novel Protein sim. GBank gi 1216486 (U48852) - HT protein [Oricetus griseus]	Contains protein domain (PF00008) - EGF-like domain	lgf	264259, 29331822, 265006, 265007, 265010, 265011, 264448, 264288, 264369, 264685, 264686, 18108357, 264768, 18108362, 264693, 18108370, 18108374, 18108379, 35696423, 83373044, 18108383, 18108385, 264584, 264585, 264587
2476	94315618 (4951, 4952)	Novel Protein sim. GBank gi 3252827 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264259, 60424269, 66714117, 264905, 265006, 264511, 265008, 265009, 264758, 265010, 265011, 18108351, 264681, 264369, 264288, 264689, 21906767, 265020, 18108374, 264639, 18108382, 83373044, 18108385, 87168518
2477	20718974 (4953, 4954)			UNCLASSIFIED	263978
2478	17659165 (4955, 4956)			UNCLASSIFIED	265017
2479	94314569 (4957, 4958)	Novel Protein sim. GBank gi 1644232[dbj BAA11082] - (D67068) N-WASP [Bos taurus]		Im7	56984075, 22278999, 21906754, 264682, 21906765
2480	95295605 (4959, 4960)			UNCLASSIFIED	264905, 264907, 264765
2481	94718481 (4961, 4962)	Novel Protein sim. GBank gi 5699469[dbj BAA83018.1] - (AB028989) KIAA1066 protein [Homo sapiens]	collagen		65274572, 56182575, 22278997, 264094, 264259, 29331822, 29331824, 66714117, 29331827, 35696052, 264508, 264905, 264906, 264907, 264908, 52644045, 264909, 56182435, 265008, 264910, 33657402, 55812038, 264758, 265010, 265011, 265017, 265018, 264760, 264762, 18108351, 264764, 264288, 264766, 264686, 264768, 21906768, 55811957, 265020, 264691, 264692, 264693, 264629, 55811576, 264630, 264634, 264635, 264636, 264637, 264556, 264558, 56182323, 83373044, 60432113, 22279002
2482	87393165 (4963, 4964)	Novel Protein sim. GBank gi 321249[pir S28407] - guanine nucleotide-exchange activator CDC25 homolog - mouse		UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 264508, 264905, 264509, 264906, 264907, 264908, 264511, 264591, 264768, 264693, 264631, 264632, 264636, 264638, 264639, 264563
2483	87731583 (4965, 4966)			UNCLASSIFIED	264488, 22278995, 264093, 264095, 60432049, 60433356, 60433438, 264448, 264288, 263967, 18108370, 18108385, 18108388, 264482
2484	94187774 (4967, 4968)	Novel Protein sim. GBank gi 728831[sp P39188]ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII		kinase	264563
2485	87786556 (4969, 4970)	Novel Protein sim. GBank gi 1185397 (U25281) - SH3 domain binding protein [Rattus norvegicus]		UNCLASSIFIED	22278995, 22278996, 22278997, 22278999, 264259, 60432049, 29331824, 60432289, 29331827, 265007, 264910, 264593, 264600, 264603, 264604, 265019, 264448, 264288, 264685, 264686, 264769, 264689, 35695917, 265022, 264692, 264693, 56182323

2486	87748978 (4971, 4972)	Novel Protein sim. GBank gi 2662167 dbj BAA23715  - (AB007803) KIAA0443 [Homo sapiens]			265017, 264555
2487	95343105 (4973, 4974)	Novel Protein sim. GBank gi 464559 sp P35287 RB14_RAT RAS-RELATED PROTEIN RAB-14	Contains protein domain (PF00071) - Ras family	glycoprotein	22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264259, 29331822, 35696052, 264108, 264905, 264907, 29331830, 264909, 265006, 264511, 265008, 265009, 60433438, 21908754, 33109954, 87168559, 265018, 264681, 264288, 264687, 21908765, 21908768, 21906767, 21908768, 21906769, 35695917, 265021, 265022, 264534, 33657023, 264692, 33657109, 263972, 18108377, 35696423, 35695855, 60170394, 18108385, 56526488, 22279000, 22279002, 264563, 264482, 264565, 20281169, 18108391
2488	87652451 (4975, 4976)			UNCLASSIFIED	264910, 264448, 264288, 264684, 264691, 264634
2489	82990585 (4977, 4978)	Novel Protein sim. GBank gi 4886439 emb CAB43355.1  - (AL050253) hypothetical protein [Homo sapiens]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complementrecept	264686, 264693, 55811576, 22279002
2490	88069609 (4979, 4980)	Novel Protein sim. GBank gi 2588624 (AC003083) - Rap2 interacting protein-like; similar to U73941 (PID:g1916018) [Homo sapiens]		UNCLASSIFIED	264907, 265008, 22279002
2491	81242116 (4981, 4982)	Novel Protein sim. GBank gi 728832 sp P39189 ALU2_HUMAN - !!! ALU SUBFAMILY SB WARNING ENTRY !!!		Im7	264259, 29331826, 265008, 264762, 18108370, 18108376, 18108379
2492	95308202 (4983, 4984)	Novel Protein sim. GBank gi 3355303 (AF001549) - Unknown gene product [Homo sapiens]		transcriptfactor	264488, 22278998, 22278999, 29331828, 264591, 33109954, 265017, 55811150, 21906764, 21906768, 264692, 60431528, 87168518, 60432113, 22279000

2493	95422415 (4985, 4986)	Novel Protein sim. GBank gi 4240307 dbj BAA74932.1  - (AB020716) KIAA0909 protein [Homo sapiens]	Contains protein domain (PF01424) - struct R3H domain	18108394, 264887, 65274572, 56182575, 22278995, 56994075, 60432049, 29331822, 29331824, 29331825, 29331826, 29331827, 29146498, 264508, 264905, 264509, 264906, 264907, 29331830, 264909, 264510, 265006, 264511, 265007, 264512, 265008, 265009, 264910, 21906754, 265011, 264600, 265017, 265018, 264604, 264605, 265019, 55811150, 264762, 18108351, 264681, 264448, 264683, 264369, 264288, 18108355, 18108357, 264687, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 264691, 264692, 33657023, 33657349, 18108370, 18108374, 18108376, 55810764, 18108379, 65274791, 264630, 264632, 264634, 264635, 264636, 264555, 264637, 264557, 264559, 264639, 264559, 83373044, 18108385, 87168518, 60432113, 22279000, 22279002, 264482, 264586, 264486
2494	30783118 (4987, 4988)			254907, 264601
2495	94234551 (4989, 4990)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED collagen 263994, 22278987, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 265008, 265009, 264595, 264604, 264446, 264682, 264764, 264288, 264685, 264766, 264769, 264689, 265020, 264692, 65274620, 264629, 55810764, 35696423, 55811576, 264636, 264637, 18108385, 22279000, 264564, 264567, 264486
2496	80018765 (4991, 4992)	Novel Protein sim. GBank gi 4808220 emb CAB42832.1  - (AL022315) dJ117715.1 (PUTATIVE novel protein) [Homo sapiens]		struct 29147620, 264905, 265006, 265007, 18108348, 18108362, 18108370, 18108374, 264555, 264556, 18108381, 18108383, 18108386
2497	91723554 (4993, 4994)			UNCLASSIFIED 52644507, 22278996, 22278999, 29331824, 29331828, 33657402, 21906754, 87168474, 265019, 264369, 264689, 21906765, 21906766, 21906767, 21906768, 265020, 33657023, 18108376, 18108387
2498	87724633 (4995, 4996)	Novel Protein sim. GBank gi 1200503 (U47924) - B [Homo sapiens]		UNCLASSIFIED 29331827, 264512, 264910, 264288, 18108374, 35695855
2499	94685125 (4997, 4998)	Novel Protein sim. GBank gi 3510234 (AC005581) - R31237.1, partial CDS [Homo sapiens]	Contains protein domain (PF000069) - kinase Eukaryotic protein kinase domain	264909, 55812038, 264631, 264637, 264558



2500	94848324 (4989, 5000)	Novel Protein sim. GBank gj3881275 emb CAA21725  - (AL032655) predicted using GeneFinder; similar to Inositol monophosphatase family; cDNA EST Yk255e11.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00459) - Inositol monophosphatase family	transport	52644507, 52645156, 22278995, 56994075, 35698286, 22278998, 264259, 52645080, 29331824, 29331825, 66714117, 60432288, 29331826, 29331827, 35696052, 29331828, 264508, 264509, 264510, 264512, 33657402, 60433438, 21906754, 52644296, 87188474, 87168559, 264603, 264681, 264448, 264683, 264288, 264369, 52644229, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 52644150, 33657023, 264693, 33657182, 35695763, 35696423, 35695855, 52644332, 83373044, 18108387, 87168518, 22279002
2501	94303896 (5001, 5002)	Novel Protein sim. GBank gj4929615 gb AAD34068.1 AF15183 - (AF151831) CGI-73 protein [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	dna_ma_bind	65274572, 56182575, 35698286, 22278996, 56994075, 22278997, 60432049, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 264905, 264906, 264907, 264908, 264909, 56182435, 264510, 264511, 265007, 264910, 264591, 60432229, 33657402, 60433356, 264595, 55812038, 264758, 264598, 87168474, 87168559, 264600, 264601, 264602, 265017, 264604, 265018, 264605, 265019, 18108351, 264448, 264369, 264288, 264766, 18108357, 21906765, 21906766, 21906767, 21906769, 26148629, 35695917, 264692, 33657023, 264629, 35695423, 55811578, 35695855, 264630, 264634, 264635, 264555, 264636, 264638, 264558, 60170394, 83373044, 18108385, 18108387, 87168518, 60432113, 22279002, 264566
2502	90993716 (5003, 5004)	Novel Protein sim. GBank gj3041847 (AC004542) - OXYSTEROL-BINDING PROTEIN-like; similar to P22059 (PID:g129308) [Homo sapiens]	Contains protein domain (PF01237) - Oxysterol-binding protein	UNCLASSIFIED	65274572, 264907, 56182435, 265007, 264592, 264760, 18108351, 264448, 264369, 264288, 264684, 264686, 55811957, 265021, 264692, 33657109, 263973, 55811576, 264635, 264555, 264556, 264557, 264558, 56182323, 264559, 87168518, 264563, 264482
2503	87878345 (5005, 5006)	Novel Protein sim. GBank gj2196874 emb CAA72638  - (Y11896) BRX protein [Mus musculus]			264905, 264907, 264512, 265008, 265011, 18108351, 264448, 264288, 29148627, 264693, 18108370, 18108374, 18108385

2504	87668706 (5007, 5008)	Novel Protein sim. GBank gij550420[emb CAA48220] - (X68101) trg [Rattus norvegicus]			264488, 52644507, 52645156, 52646842, 22278994, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264908, 264908, 52644045, 265009, 60433356, 33657402, 60433438, 264595, 33109854, 87168474, 265017, 265019, 264448, 264288, 264766, 52644229, 21906765, 21906766, 21906767, 21906768, 52644150, 264692, 27486261, 27486262, 27486264, 27486265, 35695763, 35696423, 35695855, 52644332, 56182323, 18108387, 87168518, 60432113, 22278002, 264564
2505	8760559 (5009, 5010)			UNCLASSIFIED	264605
2506	91232328 (5011, 5012)	Novel Protein sim. GBank gij2137562[pir I49635 - mouse Ohm1 protein - mouse		nuclease	264488, 52644507, 52645156, 52646365, 65274572, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264508, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 264605, 264762, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 60170615, 33657023, 27486264, 18108379, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565
2507	95316233 (5013, 5014)	Novel Protein sim. GBank gij5174489[re NP_006035.1 pKIAA - histone deacetylase 6	Contains protein domain (PF00850) - Histone deacetylase family	histone	264488, 263994, 264592, 264595, 264369, 264686, 264768, 35695917, 35696423, 264563
2508	95315505 (5015, 5016)	Novel Protein sim. GBank gij4826433[emb CAB42889.1] - (AL031447) dJ126A5.2.1 (novel protein) (isoform 1) [Homo sapiens]		UNCLASSIFIED	22278995, 22278999, 60432049, 264259, 29331828, 265006, 265007, 60433438, 33657084, 265010, 265017, 265018, 265019, 18108351, 264448, 18108354, 264369, 18108359, 21906765, 21906769, 55811957, 265020, 265022, 27486261, 33657349, 18108377, 35695855, 60432113, 22279002, 264563, 264565
2509	87813741 (5017, 5018)	Novel Protein sim. GBank gij263289 (U47656) - fibroin-4 [Araneus diadematus]		UNCLASSIFIED	265007, 265008, 18108357, 264566, 264567

2510	95421378 (5019, 5020)	Novel Protein sim. GBank gij3293537[gblAAC25762.1] - (AF071059) zinc finger RNA binding protein [Mus musculus]		dna_rna_bind	65274572, 22278994, 22278996, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 33656970, 264908, 66712502, 265007, 264910, 60170831, 60432229, 60433356, 60433438, 21906754, 87168474, 265017, 265018, 264448, 264288, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265022, 264691, 33657023, 264693, 65274620, 33657109, 33657182, 27486262, 33657349, 18108370, 35695855, 264555, 56182323, 83373044, 60432113, 22279002, 265017, 21906764, 265020, 264692
2511	87384281 (5021, 5022)	Novel Protein sim. GBank gij4323152[gblAAD162228.1] - (AF098863) Els-protein Spt-C [Mus musculus]			
2512	88084771 (5023, 5024)	Novel Protein sim. GBank gij4502075[refNP_001135.1]pAMFR - autocrine motility factor receptor	Contains protein domain (PF00087) - Zinc finger, C3HC4 type (RING finger)	transport	22278999, 264259, 29331825, 29331826, 29146499, 264907, 264909, 265006, 265008, 264591, 60432229, 21906754, 264763, 264883, 264766, 18108357, 264689, 21906769, 264693, 18108370, 263972, 18108374, 264558, 22279000
2513	95357843 (5025, 5026)	Novel Protein sim. GBank gij3004657 (AF017777) - bobby sox [Drosophila melanogaster]		UNCLASSIFIED	60424179, 52845156, 18108394, 22278994, 35696286, 56994075, 22278996, 29331822, 29331824, 60424269, 29331825, 29331827, 33656970, 60431735, 33657084, 87168559, 265017, 264448, 264369, 56181562, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 33657023, 18108366, 33657109, 27486261, 27486262, 33657349, 18108374, 55810764, 35696423, 56182323, 264558, 18108385
2514	88094578 (5027, 5028)	Novel Protein sim. GBank gij2258437 (AF008197) - syncoilin [Rattus norvegicus]		UNCLASSIFIED	264510
2515	87994509 (5029, 5030)	Novel Protein sim. GBank gij3757727[embjCAA18783] - (AL022727) dJ80119.7 olfactory receptor-like protein (hs6M1-3) [Homo sapiens]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	tm7	
2516	87786908 (5031, 5032)			UNCLASSIFIED	264259, 29146498, 264905, 264288, 29148829, 35695917, 27486281, 264634
2517	87784966 (5033, 5034)	Novel Protein sim. GBank gij4220527[embjCAA23000] - (AL035356) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	264091, 29331824, 29331825, 29331826, 29331828, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 33657402, 264757, 33109954, 265017, 265018, 264605, 264760, 264762, 264763, 264766, 264768, 264769, 33657109, 33657182, 264628, 55811576, 35696423, 264631, 264634, 264637, 264638, 264639, 87168518, 22279002, 264564

2518	94147410 (5035, 5036)	Novel Protein sim. GBank gi 4929591 gb AAD34058.1 AF15181 - (AF151819) CGI-61 protein [Homo sapiens]	Contains protein domain (PF00018) - SH3 domain	UNCLASSIFIED	35696286, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 264907, 264909, 264511, 265007, 60432229, 60433356, 60433438, 55812038, 265010, 265017, 264448, 264288, 264689, 21906768, 21906769, 265022, 52644150, 264693, 18108370, 263972, 264555, 56182323, 83373044, 18108385, 60432113, 264088
2519	94326180 (5037, 5038)	Novel Protein sim. GBank gi 4263748 gb AAD15420] - (AC004883) similar to KIAA0766; similar to PID:g3882253 [Homo sapiens]		kinase	264259, 66714117, 29331826, 29331827, 29331828, 264907, 66712502, 265008, 265008, 264594, 265010, 265011, 265018, 264288, 21906769, 265020, 60431528, 55811576, 65274791, 264632, 264555, 264636, 22279002, 264564
2520	87413235 (5039, 5040)	Novel Protein sim. GBank gi 4826722 ref NP_005085.1 pFATP - fatty acid transport protein 4		transport	264259, 264908, 264910, 264682, 21906769, 265020, 264563
2521	95316244 (5041, 5042)	Novel Protein sim. GBank gi 5174489 ref NP_006035.1 pKIAA - histone deacetylase 6	Contains protein domain (PF00850) - Histone deacetylase family	histone	264488, 264489, 263994, 65274572, 22278995, 22278998, 264259, 29331822, 29331826, 264508, 264905, 264509, 264906, 264907, 66712502, 264511, 265006, 265007, 264591, 264592, 264593, 264594, 264595, 264596, 264681, 264448, 264763, 264682, 264764, 264684, 264369, 264288, 264685, 264686, 21906768, 55811957, 264692, 264693, 27486261, 18108370, 264628, 264629, 18108374, 55811576, 35696423, 35695855, 264632, 264558, 18108385, 65274727, 60432113, 264563, 264564, 264565, 264566, 264567
2522	87754082 (5043, 5044)	Novel Protein sim. GBank gi 4580011 gb AAD24201.1 U81002_ - (U81002) TRAF4 associated factor 1 [Homo sapiens]		transport	264489, 22278997, 20281171, 21906754, 35695917, 263967, 263976, 263981, 20281169
2523	95340467 (5045, 5046)				263969
2524	95340469 (5047, 5048)	Novel Protein sim. GBank gi 1809327 (U76374) - skm- BOP2 [Mus musculus]	Contains protein domain (PF01753) - MYND finger		56984075, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 29331830, 56182435, 264512, 265008, 60170831, 33657402, 265010, 87168559, 265019, 264288, 21906765, 21906769, 35695917, 265020, 265021, 265022, 52644150, 264691, 33657023, 33657109, 27486281, 35696423, 65274791, 264559, 83373044, 56526486, 87168518, 264567

2525	94126928 (5048, 5050)	Novel Protein sim. GBank gjl2073564 (U80223) - eukaryotic initiation factor eIF-2 alpha kinase. DGCN2 [Drosophila melanogaster]		kinase	264488, 22278997, 22278999, 60432049, 60432289, 29331828, 264905, 265008, 55812038, 21906754, 265019, 264369, 21906765, 21906766, 21906767, 21906769, 35695917, 265020, 265021, 33657109, 60431528, 83373044, 60432113, 22279000, 22279002, 264565
2526	95289404 (5051, 5052)	Novel Protein sim. GBank gjl459628[d]BAA76836.1] - (AB023209) KIAA0992 protein [Homo sapiens]	Contains protein domain (PF00238) - Ribosomal protein L14	ribosomalprot	60424178, 264768, 264687, 264769, 264689, 65274572, 21906787, 56182575, 21906788, 21906789, 55811957, 22278994, 22278995, 35696286, 35695917, 22278996, 22278997, 265020, 22278998, 265021, 22278999, 265022, 264690, 264691, 60432049, 264259, 264097, 33657023, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 27486282, 264508, 264509, 264905, 264907, 18108370, 66712502, 60431528, 264828, 264909, 18108372, 18108374, 56182435, 18108376, 55810764, 55811576, 35696423, 35695855, 265006, 265007, 264512, 265008, 265009, 264634, 264635, 60431850, 264636, 264555, 264592, 60431735, 264638, 33657402, 56182323, 60433356, 60433438, 264595, 55812038, 264596, 264758, 83373044, 52646317, 18108385, 33657084, 18108387, 55811386, 65274727, 55526486, 87168518, 60432113, 265017, 22279000, 265018, 265019, 264564, 18108351, 264448, 264566, 264288, 264486, 264567, 264766
2527	88094580 (5053, 5054)	Novel Protein sim. GBank gjl2258437 (AF008197) - syncoilin [Rattus norvegicus]		UNCLASSIFIED	264510, 264512, 264630, 264591, 264592, 264259, 264594, 264595, 264603, 264605, 18108351, 264565, 264369, 18108354
2528	88078380 (5055, 5056)	Novel Protein sim. GBank gjl2085786 (AC002086) - similar to zinc finger 5 protein from Gallus gallus, U51640 (PID:g1398185) [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	dna_ma_bind	55812038, 265017, 264689, 35695917, 35695763, 60431528, 60432113, 22279002
2529	86870826 (5057, 5058)	Novel Protein sim. GBank gjl3786433 (AF098505) - similar to Arabidopsis thaliana male sterility protein 2 (SW:Q08891) [Caenorhabditis elegans]		synthase	264908, 264769, 265020, 265021, 18108383
2530	80259878 (5059, 5060)				264369, 264556
2531	87768931 (5061, 5062)			UNCLASSIFIED	29331822, 29331824, 60432289, 264508, 264509, 264906, 265011, 264769, 21906768, 33657023, 87168518, 22279000
2532	87419778 (5063, 5064)	Novel Protein sim. GBank gjl2664625[emb]CAA16972] - (AL021811) putative protein [Arabidopsis thaliana]			264553
2533	87000255 (5065, 5066)	Novel Protein sim. GBank gjl437181 (U02289) - GTPase-activating protein [Caenorhabditis elegans]		UNCLASSIFIED	264555

2534	87332322 (5067, 5068)	Novel Protein sim. GBank gi 3452473 (AF084205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]		UNCLASSIFIED	264259, 35696052, 264905, 265017, 21906769, 265020, 265022, 33657109, 22279000
2535	91225056 (5069, 5070)	Novel Protein sim. GBank gi 4468311 (emb CAB37992  - (AL031432) dJ465N24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]			65274572, 35696286, 60432289, 29331828, 66712502, 265006, 60432229, 265017, 265018, 265019, 264288, 264369, 264689, 21906768, 265020, 265021, 264636, 60170394, 22279002
2536	94218540 (5071, 5072)	Novel Protein sim. GBank gi 728838 sp P39193 ALU8_HUMAN - III! ALU SUBFAMILY SP WARNING ENTRY III	kinase		18108398, 56182575, 35696286, 22278997, 22278999, 60432049, 264259, 29331824, 29331826, 29331827, 29331828, 264905, 264511, 265009, 264910, 264596, 52646317, 18108351, 264681, 264683, 18108354, 264288, 264687, 264769, 264689, 21906765, 21906766, 21906767, 265021, 52645129, 33657109, 18108374, 18108380, 56182323, 18108381, 18108388, 87168518, 60432113, 22279000, 22279002, 264567, 18108391
2537	95422283 (5073, 5074)	Novel Protein sim. GBank gi 4557026 ref NP_003913.1 pHERC - guanine nucleotide exchange factor p532	ubiquitin		65274572, 35696286, 29331822, 29331825, 29331827, 29331828, 35696052, 264906, 66712502, 264909, 265008, 265011, 264760, 264288, 264685, 35695917, 60170615, 264691, 33657023, 65274620, 33657109, 18108374, 35696423, 35695855, 264636, 264558, 60170394, 56182323, 83373044
2538	36853454 (5075, 5076)			UNCLASSIFIED	
2539	9414916 (5077, 5078)			UNCLASSIFIED	22278996, 22278999, 29331822, 29331825, 29331828, 29146499, 264908, 264112, 60170831, 87168559, 264604, 265019, 264685, 264766, 87168518, 22279000, 264565, 264566
2540	94218545 (5079, 5080)	Novel Protein sim. GBank gi 1362647 pir J53876 - sex-regulated protein janus A - fruit fly (Drosophila pseudoobscura)		UNCLASSIFIED	22278997, 29331828, 265006, 265009, 264758, 265010, 18108351, 264683, 264288, 21906765, 35695917, 265020, 18108374, 264567
2541	95308238 (5081, 5082)	Novel Protein sim. GBank gi 1711658 sp P54787 T10_MOUSE - SER/THR-RICH PROTEIN T10 IN DGCR REGION		UNCLASSIFIED	264488, 264768, 264689, 264511, 20281171, 264634, 264635, 264691, 264639, 29331824, 264603, 264604, 264905, 264907, 264908, 264768

2542	95298182 (5083, 5084)	Novel Protein sim. GBank gi 5225320 gb AAD40850.1 AF08310 - (AF083107) sirtuin type 2 [Homo sapiens]	Contains protein domain (PF00220) - Neurohypophyseal hormones, N- terminal Domain	UNCLASSIFIED	264488, 18108394, 52646365, 52646842, 65274572, 22278994, 35698286, 22278996, 264259, 52645080, 29331822, 29331824, 29331827, 35696052, 33658970, 264907, 264909, 52644045, 264510, 265008, 284512, 265007, 265008, 265009, 264910, 60431735, 52646317, 52644296, 265010, 265011, 265018, 265019, 18108351, 264683, 264288, 264685, 264687, 52644229, 264769, 21906768, 21906767, 21906769, 52644150, 33657023, 33657109, 52645129, 33657182, 27488261, 27488264, 33657349, 35695763, 18108374, 35696423, 35695855, 264631, 264634, 264635, 264558, 83373044, 18108385, 18108387, 87168518, 264563, 264584
2543	94139088 (5085, 5086)	Novel Protein sim. GBank gi 5419857 emb CAB46374.1  - (AL096723) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)		65274572, 56182575, 22278999, 264259, 29331826, 264907, 264510, 264511, 264592, 264595, 264764, 264369, 264288, 264684, 264766, 264689, 21906765, 21906767, 21906769, 60170615, 264692, 264693, 55811576, 65274791, 264636, 264556, 18108381, 60170394, 264639, 18108385, 60432113, 22279000
2544	94218548 (5087, 5088)	Novel Protein sim. GBank gi 2498110 sp Q63191 AEGP_RAT - APICAL ENDOSOMAL GLYCOPROTEIN PRECURSOR	Contains protein domain (PF00629) - MAM domain.	glycoprotein	18108397, 52646365, 22278997, 264259, 60432049, 29331822, 29331825, 29331826, 29331827, 29331828, 264905, 264908, 265006, 265007, 265008, 87168559, 265017, 265018, 265019, 18108351, 264448, 264686, 264687, 264689, 21906765, 265020, 265021, 18108370, 18108374, 18108376, 18108381, 18108385, 18108387, 56526486, 22278000, 264482, 264563, 264567
2545	87742845 (5089, 5090)	Novel Protein sim. GBank gi 3327046 dbj BAA31591  - (AB014516) KIAA0616 protein [Homo sapiens]			29331825, 264906, 265009, 60170831, 265017, 264369, 21906767, 60170815, 264692, 33657109
2546	88093861 (5091, 5092)	Novel Protein sim. GBank gi 2996032 AF054586  - brain finger protein [Rattus norvegicus]	Contains protein domain (PF00087) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	29331824, 265007, 22278002

2547	94143868 (5093, 5094)	Novel Protein sim. GBank gi 4929607 gb AAD34064.1 AF15182 - (AF151827) CGI-69 protein [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	- transport	264488, 18108394, 52646842, 18108397, 56182575, 22278995, 56994075, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 35696052, 29331828, 264104, 264508, 264905, 264906, 264908, 66712502, 264909, 56182435, 265006, 265007, 264512, 265008, 265009, 60170831, 60432229, 60431735, 264594, 60433438, 21906754, 52646317, 265010, 265011, 264600, 264601, 265018, 265019, 264760, 18108351, 264682, 264448, 264288, 264369, 264684, 264686, 264687, 56181562, 264688, 264689, 21906785, 21906766, 21906767, 21906768, 29148627, 21906769, 55811957, 265020, 265021, 265022, 264690, 264691, 18108362, 264692, 264693, 27486261, 18108370, 18108374, 55810764, 55811576, 35696423, 35695855, 264635, 264636, 264555, 264637, 263981, 264557, 18108380, 264638, 56182323, 264558, 264559, 83373044, 18108385, 87168518, 22279002, 264584, 264586, 264486
2548	88179079 (5095, 5096)			UNCLASSIFIED	264488, 18108394, 52646365, 22278994, 35696286, 56994075, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 56182435, 264511, 265007, 264512, 60433356, 87168559, 264684, 264369, 52644229, 265021, 33657023, 264692, 18108374, 52644332, 264557, 18108380, 18108381, 18108382, 18108384, 18108385, 60432113, 22279000, 22279002, 264563, 264567
2549	94196893 (5097, 5098)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII	Contains protein domain (PF00412) - LIM domain containing proteins	- struct	56182575, 22278996, 22278997, 22278998, 22278999, 264259, 264508, 264908, 29331830, 265009, 265010, 265018, 264688, 21906764, 21906765, 21906766, 21908767, 21906769, 265020, 265021, 52644150, 264691, 18108368, 60431602, 18108376, 35696423, 56182323, 18108387, 264567, 56182575, 29331822, 264105, 264512, 18108351, 35695917, 264637, 264638
2550	87778584 (5099, 5100)	Novel Protein sim. GBank gi 2143886 pr I 52523 - nucleoporin p62 homolog - rat (fragment)		UNCLASSIFIED	



2551	95308400 (5101, 5102)	Novel Protein sim. GBank gi 4337103 gb AAD18079  - (AF129756) NG26 [Homo sapiens]	Contains protein domain (PF00561) - alpha/beta hydrolase fold	UNCLASSIFIED	18108396, 65274572, 22278995, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 264905, 56182435, 265007, 60433438, 55812038, 21906754, 65274444, 265017, 265018, 264605, 265019, 264288, 21906766, 21906768, 21906769, 265020, 60170615, 264693, 33657109, 35696423, 264638, 56182323, 83373044, 22279000
2552	95332620 (5103, 5104)			UNCLASSIFIED	56182575, 35696286, 29331824, 29331826, 35696052, 29331828, 264508, 264907, 56182435, 265008, 264591, 33109954, 264760, 55811957, 35695917, 33657023, 33657109, 18108374, 55811576, 35696423, 35695855, 56182323, 264558
2553	95308243 (5105, 5106)	Novel Protein sim. GBank gi 171658 sp P54797 T10_MOUSE - SER/THR-RICH PROTEIN T10 IN DGCR REGION		UNCLASSIFIED	264686, 264488, 263976, 264768, 29331826, 35696052, 35696423, 264601, 264511, 264602, 264910, 264634, 264760, 264555, 264762, 264906, 264592, 264691, 264566, 264908, 264684, 264567, 264909, 264766
2554	87761520 (5107, 5108)	Novel Protein sim. GBank gi 728835 sp P38192 ALU5_HUMAN - IIII ALU SUBFAMILY SC WARNING ENTRY IIII		cadherin	22278997, 29331822, 264508, 21908769, 33657023, 33657109, 56182323
2555	87627551 (5109, 5110)	Novel Protein sim. GBank gi 4884319 emb CAB43260.1  - (AL050084) hypothetical protein [Homo sapiens]		nuclease	29331824, 263872
2556	87645533 (5111, 5112)	Novel Protein sim. GBank gi 4106984 (AC003038) - R30923_1 [Homo sapiens]	Contains protein domain (PF00514) - Armadillo/beta-catenin-like repeats	UNCLASSIFIED	22278998, 264509, 33657402, 264683, 264684, 264766, 264689, 33657023, 33657109, 35695855, 264558, 264567, 264595
2557	79437803 (5113, 5114)	Novel Protein sim. GBank gi 119110 sp P03211 EBN1_EBV - EBNA-1 NUCLEAR PROTEIN		UNCLASSIFIED	22278997, 29331824, 66714117, 29331825, 264906, 264511, 265018, 264448
2559	88096382 (5117, 5118)	Novel Protein sim. GBank gi 4538998 emb CAB39619.1  - (AL049481) AIG1-like protein [Arabidopsis thaliana]		UNCLASSIFIED	22278997, 29331822, 29331828, 60433356, 265011, 264288, 264765, 264766, 264769, 21906765, 21906766, 60432113, 264482
2560	87994530 (5119, 5120)	Novel Protein sim. GBank gi 5051399 emb CAB44995.1  - (AL078630) 573K1.3 (mm17M1-4) (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor LIKE) protein) [Mus musculus]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)		
2561	88176575 (5121, 5122)	Novel Protein sim. GBank gi 5326825 gb AAD42056.1 AF044953 - (AF044953) NADH:ubiquinone oxidoreductase PGIV subunit [Homo sapiens]		UNCLASSIFIED	22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 265007, 60432228, 87168559, 265017, 265018, 265019, 264689, 21906766, 21906769, 35695917, 265020, 33657023, 33657109, 18108374, 264634, 264559, 18108385, 87168518, 22279002

2562	87645539 (5123, 5124)	Novel Protein sim. GBank gi 4106984 (AC003038) - R30923_1 [Homo sapiens]		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433356, 21908765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518
2563	88095497 (5125, 5126)	Novel Protein sim. GBank gi 4886447 emb CAB43371.1  - (AL050270) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486262, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108384, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2564	80502783 (5127, 5128)	Novel Protein sim. GBank gi 1352944 sp P47179 Y9P_YEAST - HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR		sulfoltransferase	29331822, 265007, 265010, 265019, 264769, 55811576, 56182323
2565	85530905 (5129, 5130)			UNCLASSIFIED	68714117, 284909, 263978, 264632
2566	80224956 (5131, 5132)	Novel Protein sim. GBank gi 628012 pir A53933 - myosin I myr 4 - rat	Contains protein domain (PF000083) - striated Myosin head (motor domain)		18108370, 35695855, 264556, 264558, 18108383
2567	86143590 (5133, 5134)				265020, 60170615
2568	91233099 (5135, 5136)	Novel Protein sim. GBank gi 466009 sp P34548 YNJ4_CAEEL - HYPOTHETICAL 33.8 KD PROTEIN R10E11.4 IN CHROMOSOME III		synthase	60424179, 18108394, 56181686, 56994075, 22278999, 264490, 264259, 29331822, 56182181, 29331824, 60424269, 29331825, 29331826, 29331828, 264509, 29331830, 265007, 265008, 265009, 33657402, 265010, 265011, 265019, 264448, 264683, 264288, 18108354, 264769, 21906766, 21906767, 35695917, 265021, 33657023, 18108362, 33657109, 33657182, 35695763, 60431528, 55810764, 18108379, 83373044, 18108385, 60432113, 264482

2568	95313764 (5137, 5138)	Novel Protein sim. GBank gi 259560 gb AABB4166.1  - (AF029674) basic leucine zipper protein LZIP [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	transcriptfactor	18108394, 56182575, 56181686, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 56182181, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264905, 264906, 56182435, 265008, 265009, 264910, 60432229, 264592, 60433356, 60433438, 21906754, 87168559, 265017, 265018, 265019, 264682, 264448, 264288, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 35695917, 265021, 265022, 52644150, 264690, 264691, 264692, 264693, 65274620, 263967, 35695763, 20281069, 263974, 18108374, 55810764, 35698423, 35695855, 264558, 18108381, 56182323, 18108382, 83373044, 18108385, 56526486, 87168518, 22279000, 22279002, 264566
2570	94136764 (5139, 5140)	Novel Protein sim. GBank gi 4758954 ref NP_004567.1 pPPP2 - protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), beta isoform	Contains protein domain (PF01240) - Protein phosphatase 2A regulatory subunit PR55	phosphatase	22278998, 29331822, 29331824, 66714117, 29331825, 60432289, 29331827, 35696052, 264907, 264510, 265007, 265009, 264758, 33109854, 265019, 264686, 264689, 265020, 265021, 264691, 264692, 264693, 18108374, 35695855, 264634, 56182323, 264639, 60170394, 83373044, 22279002, 264482
2571	87733760 (5141, 5142)	Novel Protein sim. GBank gi 732218 sp P34609 YO60_CAEEL - HYPOTHETICAL 128.6 KD PROTEIN ZK1098.10 IN CHROMOSOME III		strud	264508, 264905, 264907, 264828, 18108351, 264555, 264556, 264557, 264558, 264559
2572	87627560 (5143, 5144)	Novel Protein sim. GBank gi 4894319 emb CAB43260.1  - (AL050084) hypothetical protein [Homo sapiens]		UNCLASSIFIED	35696286, 29331827, 35696052, 264100, 264104, 264110, 264592, 21905754, 29148627, 29148629, 263972, 263974, 18108374, 263976, 35695855, 60170394, 264559, 18108385

2573	85313928 (5145, 5146)	Novel Protein sim. GBank gi 398138 sp P02745 C1QA_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, A CHAIN PRECURSOR	Contains protein domain (PF00336) - C1q domain	complement	264488, 60424179, 65274572, 56182575, 56181686, 22278995, 56994075, 22278997, 60432049, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 264104, 264107, 264508, 264906, 29331830, 264909, 264510, 265008, 264512, 265008, 265009, 264910, 264591, 264592, 60432229, 264593, 60433356, 264594, 60433438, 264595, 55812038, 264759, 21906754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264761, 264762, 264763, 264764, 264389, 264288, 264685, 264766, 264686, 264687, 264688, 264769, 56181562, 264689, 21906765, 21906766, 21906767, 29148627, 21906768, 21906769, 265020, 265021, 265022, 60170615, 264690, 52644150, 264691, 264692, 33657023, 65274620, 18108365, 18108368, 27486285, 60431802, 264629, 60431528, 263976, 65274791, 35695855, 20281071, 60431850, 264637, 264638, 264558, 264639, 56182323, 60170394, 83373044, 18108384, 87168518, 60432113, 264482, 264564, 264585, 264566, 264567
2574	94746814 (5147, 5148)	Novel Protein sim. GBank gi 3334982 (AC005306) - R27216_1 [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	UNCLASSIFIED	22278995, 264259, 60432289, 29331827, 29331828, 33656970, 264908, 265008, 264910, 264591, 33657402, 265018, 265019, 264448, 264764, 264389, 264288, 18108357, 21906765, 21906766, 21906768, 55811957, 60170615, 264691, 33657023, 264693, 33657109, 33657182, 27486261, 27486264, 33657349, 264636, 264555, 83373044, 18108385, 264482
2575	87754408 (5149, 5150)	Novel Protein sim. GBank gi 4929729 gb AAD34125.1 AF15188 - (AF15188) CGI-130 protein [Homo sapiens]		UNCLASSIFIED	264910, 264601, 264683, 264689, 264080
2576	95357881 (5151, 5152)	Novel Protein sim. GBank gi 4680681 gb AAD27730.1 AF132955 - (AF132955) CGI-21 protein [Homo sapiens]	Contains protein domain (PF00442) - Ubiquitin carboxyl-terminal hydrolases family 2	ubiquitin	264259, 35696052, 264906, 60433438, 264681, 18108351, 264288, 52644150, 264628, 35696423
2577	86996621 (5153, 5154)	Novel Protein sim. GBank gi 4337103 gb AAD18079  - (AF129756) NG26 [Homo sapiens]		UNCLASSIFIED	29331825, 265018, 265019, 264685
2578	87786941 (5155, 5156)			UNCLASSIFIED	264488, 264906, 264908, 264910, 264596, 264603, 264604, 264605, 264766, 21906769, 264628, 264630, 264634, 264639, 264563
2579	87292879 (5157, 5158)			UNCLASSIFIED	29331822, 29331824, 264767

2580	88166788 (5159, 5160)	Novel Protein sim. GBank gi 2588628 (AC003080) - Similar to KIAA0299; 60% similarity to AB002297 (PID:g2224539) [Homo sapiens]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).		265007, 265018, 264762
2581	87899048 (5181, 5182)	Novel Protein sim. GBank gi 4406642 gb AAD20048  - (AF131809) Unknown [Homo sapiens]		collagen	56994075, 29331824, 29331826, 29331828, 264805, 60433358, 60433438, 264758, 87168559, 21906769, 265022, 35695855, 263981
2582	87786788 (5163, 5164)	Novel Protein sim. GBank gi 2739387 (AC002505) - putative phosphatidylinositol-4-phosphate 5-kinase [Arabidopsis thaliana]		eph	264488, 264907, 264908, 264910, 264764, 264684, 264766, 264836, 264555, 264565
2583	91220850 (5165, 5166)	Novel Protein sim. GBank gi 4378112 emb CAA16521.1  - (AL021578) dJ453C12.2 (similar to transcription factor RBP-L) [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	- transcript factor	56181686, 264259, 264510, 264512, 264591, 264592, 264593, 264594, 264595, 264596, 264603, 264629, 55810764, 264630, 264637, 264565
2584	80430941 (5167, 5168)			UNCLASSIFIED	264908, 264910, 264768, 264693, 18108374, 55811576, 56182323
2585	80436126 (5169, 5170)	Novel Protein sim. GBank gi 2736151 (AF021935) - myotonic dystrophy kinase-related Cdc42-binding kinase [Rattus norvegicus]		kinase	264768
2586	91226136 (5171, 5172)				22278998, 264259, 29331822, 29331824, 29331827, 29331828, 264906, 265007, 265009, 264591, 60433356, 33657402, 265018, 264762, 264288, 21906766, 21906767, 21906769, 265022, 264691, 83373044, 56526486, 22279002
2587	80430943 (5173, 5174)				264908, 265018, 264768, 264693, 55811576, 56182323
2588	80074385 (5175, 5176)			UNCLASSIFIED	264564
2589	85515607 (5177, 5178)	Novel Protein sim. GBank gi 3021598 emb CAA71415  - (Y10389) nuclear protein [Xenopus laevis]		UNCLASSIFIED	35696052, 264905, 264906, 264907, 264908, 264908, 265009, 265018, 264769, 35698423, 264636
2590	87054526 (5179, 5180)	Novel Protein sim. GBank gi 2104689 (U92793) - alpha glucosidase II, alpha subunit [Mus musculus]	Contains protein domain (PF01055) - Glycosyl hydrolases family 31	glucoamylase	22278995, 29331830, 265008, 265010, 265017, 264639
2591	94192167 (5181, 5182)	Novel Protein sim. GBank gi 5702202 gb AAD47199.1 AF12916 - (AF129166) long-chain acyl-CoA synthetase 5 [Homo sapiens]		eph	264259, 29331822, 264106, 264906, 56182435, 265007, 265008, 33109954, 264448, 55811957, 265020, 18108370, 55811576, 22279002

2592	95332648 (5183, 5184)	Novel Protein sim. GBank gi 3024988 sp Q60936 YAB1_MOUSE - HYPOTHETICAL HEART PROTEIN	transport	18108397, 56182575, 35696286, 56994075, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 29331828, 264906, 264908, 265007, 265008, 264910, 60432229, 264594, 60433356, 60433438, 55812038, 18108348, 21906754, 265011, 87168559, 265017, 265019, 264764, 264369, 264288, 264766, 265021, 60170815, 33657023, 33657109, 264629, 35696423, 35695855, 264557, 264638, 60170394, 56182323, 83373044, 56526486, 87168518, 264563, 264482, 264565
2593	87754416 (5185, 5186)	Novel Protein sim. GBank gi 4929729 gb AAD34125.1 AF15188 - (AF151888) CGI-130 protein [Homo sapiens]	tm7	22278999, 29331825, 264758, 21906754, 52646317, 265010, 18108351, 264288, 264369, 21906768, 264693, 18108370, 264637, 264638, 264482
2594	95305758 (5187, 5188)	Novel Protein sim. GBank gi 4929587 gb AAD34054.1 AF15181 - (AF151817) CGI-59 protein [Homo sapiens]	UNCLASSIFIED	264488, 18108398, 56182575, 35696286, 22278997, 264093, 264259, 29331822, 29331825, 68714117, 29331826, 264905, 264909, 52644045, 56182435, 264510, 264512, 265007, 264757, 21906754, 87168474, 265017, 264760, 264448, 264764, 264288, 264766, 264689, 21906768, 33657109, 263975, 263977, 264634, 264556, 60170394, 56182323, 56526486, 264482, 264563, 264564, 264586, 264587
2595	79561676 (5189, 5190)	Novel Protein sim. GBank gi 4309681 gb AAD15478  - (AC006930) R33423_1 [Homo sapiens]	UNCLASSIFIED	264692
2596	87538637 (5191, 5192)		UNCLASSIFIED	22278999, 264259, 265018, 264448, 265021, 60431528
2597	94784089 (5193, 5194)		UNCLASSIFIED	264905, 264509, 264908, 264762, 264766, 35695917, 35695855, 264635, 264636, 83373044, 264486
2598	88094948 (5195, 5196)	Novel Protein sim. GBank gi 1001351 db BAA10838  - (D84006) hypothetical protein [Synecchocystis sp.]	UNCLASSIFIED	22278998, 264259, 29331824, 87168474, 264683, 21906766, 35695917, 264691, 33657023, 33657109, 18108370, 18108374, 264564, 264585
2598	87842889 (5197, 5198)	Novel Protein sim. GBank gi 3941737 (AF109719) - BAT2 [Mus musculus]	MHC	264766, 264769, 21906768, 33657182, 35695763, 18108370, 18108374, 264635, 264636, 56526486, 22278900, 264586
2600	87787846 (5199, 5200)	Novel Protein sim. GBank gi 4263521 gb AAD15347  - (AC004044) putative WD-repeat protein [Arabidopsis thaliana]	kinasereceptor	35696286, 264093, 264288, 21906769, 35696423, 35695855

2601	91243070 (5201, 5202)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		kinase	5618257, 22278999, 264259, 29331822, 29331824, 60432289, 29331826, 29331827, 264806, 265007, 265008, 264591, 60433356, 33657402, 60433438, 21906754, 265011, 265018, 265019, 18108351, 264448, 264369, 21906769, 265020, 60170615, 264693, 33857109, 18108370, 18108376, 58182323, 18108381, 18108385, 22279002, 264563 60433438, 21906754, 87168559, 264601, 264369, 264288, 21906767
2602	88180022 (5203, 5204)	Novel Protein sim. GBank gi 406632 gb AAD200471 - (AF131801) Unknown [Homo sapiens]			
2603	94325821 (5205, 5206)	Novel Protein sim. GBank gi 3122367 sp O61211 LIGA_MOUSE - LIGATIN		UNCLASSIFIED	264488, 65274572, 22278995, 22278996, 56984075, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35696052, 56182435, 264113, 265008, 265009, 60433356, 264757, 60433438, 264759, 33657084, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264683, 18108354, 264288, 264767, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486262, 18108374, 35696423, 65274791, 35695855, 264555, 264636, 264637, 58182323, 83373044, 56526486, 87168518, 60432113, 22279000
2604	94676601 (5207, 5208)	Novel Protein sim. GBank gi 5454030 ref NP_006488.1 pRRP2 - RAS-related on chromosome 22		oncogene	264259, 35696052, 264508, 264906, 264907, 264908, 264909, 264510, 264512, 265008, 264810, 33657402, 264804, 264605, 264762, 264763, 264682, 264764, 264683, 264768, 264769, 264688, 33657023, 264693, 18108365, 264628, 35696423, 264631, 264632, 264634, 264635, 264637, 18108381, 264639, 83373044, 264565
2605	94316766 (5209, 5210)	Novel Protein sim. GBank gi 3628745 dbj BAA33366 - (AB013721) milisugumin 23 [Oryctolagus cuniculus]		UNCLASSIFIED	22278998, 264490, 60432049, 264259, 60432289, 264909, 265008, 60433356, 60433438, 264758, 21906754, 265010, 265011, 265018, 264681, 18108351, 264288, 264766, 264685, 21906765, 21906766, 21906768, 21906769, 264691, 264692, 264693, 65274791, 264634, 264555, 264636 22278996, 264510, 264512, 265009, 264766, 22279002, 264566
2606	87746406 (5211, 5212)				

2607	87627742 (5213, 5214)	Novel Protein sim. GBank gi14826626[gb AAO30202.1] - (AF135022) mediator [Homo sapiens]			29331822, 29331825, 29331826, 29331827, 29331828, 264906, 264907, 264908, 66712502, 264828, 56182435, 55812038, 265010, 265017, 265018, 265019, 264768, 264689, 21906765, 55811957, 265020, 265022, 264692, 33657023, 264693, 33657109, 18108370, 264639, 56182323, 264509, 264907, 264908, 264592, 264758, 264631
2608	81734786 (5215, 5216)	Novel Protein sim. GBank gi12226005 (U49973) - ORF2: function unknown [Homo sapiens]			
2609	94843791 (5217, 5218)	Novel Protein sim. GBank gi13024889[sp P56524 Y288_HUMAN - HYPOTHETICAL PROTEIN KIAA0288 (HA6116)]	Contains protein domain (PF00850) - Histone deacetylase family	histone	264488, 65274572, 35696286, 22278997, 22278999, 60432049, 264259, 56182181, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264905, 264907, 60433356, 60433438, 55812038, 265011, 87168559, 265017, 265018, 264448, 264765, 264288, 264766, 264689, 21908765, 21908767, 21906769, 265020, 265021, 264691, 264692, 33657109, 27486261, 18108370, 65274791, 264638, 264556, 56182323, 18108385, 56526486
2610	38177684 (5219, 5220)	Novel Protein sim. GBank gi14336855[gb AAD17989] - (AF108473) leucine-rich-domain inter-acting protein 1; LeR inter-acting protein 1; LEAP1 [Mus musculus]		transcript factor	18108394, 22278994, 56994075, 60432049, 264259, 29331822, 29331825, 60432289, 29331827, 264107, 264109, 264905, 56182435, 284112, 265006, 265007, 265008, 265009, 60433356, 60433438, 265011, 87168559, 265017, 264448, 264682, 264764, 264288, 265021, 33657023, 263967, 33657182, 27486261, 18108374, 263976, 55811576, 264638, 87168518, 60432113, 22278999, 265017, 264684, 21908768, 22279000
2611	87428890 (5221, 5222)	Novel Protein sim. GBank gi13876761[emb CAA92994] - (Z68760) predicted using GeneFinder; Similarity to Mouse FK506-binding protein (SW:FKB3_MOUSE) [Caenorhabditis elegans]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	
2612	87771188 (5223, 5224)	Novel Protein sim. GBank gi15679136[gb AAD46874.1 AF16093 - (AF160934) BcDNA.LD14189 [Drosophila melanogaster]		transport	265009, 264910, 264759, 265017, 21908767, 18108365, 18108388, 60432113
2613	79481486 (5225, 5226)	Novel Protein sim. GBank gi15679136[gb AAD46874.1 AF16093 - (AF160934) BcDNA.LD14189 [Drosophila melanogaster]		UNCLASSIFIED	264685
2614	87643948 (5227, 5228)	Novel Protein sim. GBank gi15533081[gb AAD45009.1 AF16118 - (AF161181) P55T protein [Mus musculus]	Contains protein domain (PF00625) - Guanylate kinase		22278998, 22278999, 29331825, 264508, 264906, 21908754, 264602, 264766, 264769, 52644229, 21908765, 33657109, 27488264, 18108370, 263972, 264555, 60432113
2615	87381896 (5229, 5230)			UNCLASSIFIED	264768, 18108394, 264692, 264693, 264508, 264509, 264907, 264628, 264908, 264909, 18108377, 264511, 264512, 264910, 264635, 264595, 265010, 264404, 264563, 264764, 264685, 264768



2616	87428895 (5231, 5232)	Novel Protein sim. GBank gl 3876761 emb CAA929941 - (Z68760) predicted using GeneFinder; Similarity to Mouse FK506-binding protein (SW:FKB3_MOUSE) [Caenorhabditis elegans]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	22278995, 22278997, 22278998, 60432049, 60432289, 264828, 60433356, 264594, 60433438, 33109954, 87168474, 265011, 265017, 265019, 264288, 264766, 21906765, 21906767, 265020, 265021, 18108376, 18108377, 18108387, 87168518, 264482, 264567
2617	86976888 (5233, 5234)	Novel Protein sim. GBank gl 728831 sp P39186 ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII		kinase	265010, 265019, 264369, 264693, 55811576, 22278002
2618	91231662 (5235, 5236)	Novel Protein sim. GBank gl 3319282 (AF049103) - Huntingtin interacting protein [Homo sapiens]	Contains protein domain (PF00397) - WW domain	UNCLASSIFIED	264489, 22278996, 264490, 264259, 29331822, 264102, 264509, 264906, 264907, 68712502, 29331830, 265008, 264910, 265009, 60433356, 60433438, 264758, 21906754, 265011, 87168559, 265017, 265018, 264369, 264288, 264766, 264768, 264689, 21906765, 21906766, 21906767, 35695917, 265020, 265022, 33657023, 264692, 33657109, 264628, 18108374, 35695855, 18108381, 83373044, 18108385, 18108388, 56526486, 264563
2619	87694000 (5237, 5238)	Novel Protein sim. GBank gl 2431772 (U66411) - putative type III alcohol dehydrogenase [Drosophila melanogaster]	Contains protein domain (PF00465) - Iron-containing alcohol dehydrogenases	dehydrogenase	264259, 60432289, 60433438, 21906754, 264369, 60432113, 264566
2620	95314841 (5239, 5240)	Novel Protein sim. GBank gl 4322567 gb AAD160971 - (AF090436) dachshund variant 1 [Mus musculus]	Contains protein domain (PF00628) - PHD-finger	UNCLASSIFIED	52644507, 52645156, 52646842, 65274572, 22278995, 56994075, 35696286, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331828, 35696052, 264907, 68712502, 265008, 60433356, 33657402, 52646317, 21906754, 87168474, 265010, 265017, 265018, 264448, 264369, 264288, 264687, 264768, 52644229, 264688, 264689, 21906765, 21906766, 35695917, 52644150, 264692, 33657109, 35695763, 35696423, 264556, 52644332, 18108382, 83373044, 18108385, 18108387, 65274727, 87168518, 60432113, 22278002, 264594, 264636
2621	80253495 (5241, 5242)				
2622	81780390 (5243, 5244)	Novel Protein sim. GBank gl 4557341 ref NP_001174.1 pATP6 - ATPase, H <sup>+</sup> transporting, lysosomal subunit 1; vacuolar proton pump, H-ATPase subunit			264488, 264906, 264907, 264908, 264512, 265007, 264758, 35695917, 264634, 264636, 264563, 264482
2623	91639306 (5245, 5246)	Novel Protein sim. GBank gl 3880355 emb CAB052991 - (Z82285) predicted using GeneFinder [Caenorhabditis elegans]		UNCLASSIFIED	29331824, 35696052, 265007, 265010, 264288, 29148629

2624	91639308 (5247, 5248)	Novel Protein sim. GBank gij3880355[embjCAB05299] - (Z82285) predicted using GeneFinder [Caenorhabditis elegans]		UNCLASSIFIED	56181686, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 56182181, 29331825, 60432289, 29331828, 35696052, 29146489, 66712502, 52644045, 265007, 265008, 60433356, 33109954, 21908754, 265010, 265011, 265018, 264448, 264288, 21908765, 21908766, 21908767, 29148629, 35695917, 265021, 265022, 27486265, 18108370, 60431528, 55811576, 35695855, 56182323, 18108385, 87168518, 22279002, 18108391
2625	86452068 (5249, 5250)	Novel Protein sim. GBank gij2887429[dbjBAA24857] - (AB007887) KIAA0427 [Homo sapiens]		UNCLASSIFIED	264091, 264511, 263981
2626	16533787 (5251, 5252)	Novel Protein sim. GBank gij487416 (L20302) - actin filament protein [Gallus gallus]		struct	265008
2627	87636823 (5253, 5254)	Novel Protein sim. GBank gij88462[pirjA27307] - proline-rich phosphoprotein (gene PRH1, Db allele) - human		UNCLASSIFIED	22278996, 265007, 265009, 264448, 21908767, 265021, 264558, 87168518
2628	94848254 (5255, 5256)	Novel Protein sim. GBank gij3123552[embjCAA18609] - (AL022578) dJ393P12.2 (hypothetical Proline-rich protein KIAA0269 LIKE) [Homo sapiens]		UNCLASSIFIED	22278997, 22278999, 264259, 60432049, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 264907, 264909, 265008, 264591, 60433356, 60433438, 265010, 265017, 265018, 264369, 264288, 18108357, 21908765, 21908768, 265022, 65274791, 264638, 18108387, 87168518, 22279002
2629	87376490 (5257, 5258)	Novel Protein sim. GBank gij4929595[gbjAAD34058.1]AF15182 - (AF151821) CGI-63 protein [Homo sapiens]		synthase	29331825, 29331826, 264102, 265006, 264766, 35695917, 264691, 33657023, 263972, 18108374, 22279000
2630	78188364 (5259, 5260)			UNCLASSIFIED	264636, 18108385
2631	94845909 (5261, 5262)	Novel Protein sim. GBank gij321605[pirjJQ1161] - Gag protein - Visna virus (strain EV1)	Contains protein domain (PF00098) - Zinc finger, CCHC class	UNCLASSIFIED - dna_ma_bind	52644507, 52645156, 52646365, 52646842, 22278994, 22278995, 35696288, 56994075, 22278997, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 33656970, 264905, 264509, 264907, 264908, 264511, 264512, 265007, 265008, 264910, 52646317, 33657084, 52644296, 265010, 87168559, 265017, 265018, 265019, 264760, 264762, 264448, 264288, 264369, 264766, 264768, 52644229, 21908764, 21908765, 21908766, 21908767, 21908769, 35695917, 265020, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486265, 33657349, 35695763, 35696423, 65274791, 35695855, 264634, 264637, 52644332, 56182323, 60432113, 264566, 264486, 264685
2632	38730414 (5263, 5264)				

2633	95011617 (5265, 5266)	Novel Protein sim. GBank gij113954[db][BAA10889] - (D84009) seizure-related gene product 6 type 2 precursor [Mus musculus]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	22278995, 22278997, 22278998, 264259, 29331822, 29331824, 29331825, 29331827, 264508, 264906, 265006, 265007, 265008, 265009, 55812038, 33657084, 55811386, 265010, 265011, 87168559, 265018, 265019, 264683, 264288, 264686, 29148629, 33657023, 264693, 33657182, 35695763, 55811576, 264639, 56182323, 83373044, 18108385, 56526486, 87168518, 22279000, 22279002, 264565
2634	87330921 (5267, 5268)	Novel Protein sim. GBank gij5441611[emb][CAB46854.1] - (AJ388555) hypothetical protein [Canis familiaris]	UNCLASSIFIED	29331826, 263972, 264089
2635	86623144 (5269, 5270)	Novel Protein sim. GBank gij4680663[gb][AAD27721.1]AF13294 - (AF132946) CGI-12 protein [Homo sapiens]		22278997, 264259, 29331824, 66714117, 29331827, 29331828, 264907, 33657084, 265017, 265018, 264448, 264288, 21906766, 21906767, 21906768, 29148629, 18108376, 55811576, 35695855, 87168518, 22279000
2636	87260534 (5271, 5272)	Novel Protein sim. GBank gij3879146[emb][CAB07646] - (Z93386) Similarity to Yeast hypothetical 52.9 KD protein (SW:P43816); cDNA EST EMBL:M89432 comes from this gene; cDNA EST EMBL:D71008 comes from this gene; cDNA EST EMBL:D73578 comes from this gene; cDNA EST EMBL:D69025 comes ...	Contains protein domain (PF01546) - Peptidase family M20/M25/M40	264569, 29331822, 29331828, 265008, 60170831, 264681, 264765, 264685, 29148627, 21906769, 29148784, 265022, 60170815, 264635, 18108385, 56526486, 22279002, 264567
2637	95011298 (5273, 5274)	Novel Protein sim. GBank gij4758208[ref][NP_004081.1]pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	264488, 264489, 52644507, 264887, 52646365, 52646842, 22278994, 22278996, 22278998, 20281171, 264259, 29331822, 52645080, 66714117, 29331825, 29331826, 29331827, 35696052, 29331828, 29146488, 264508, 264905, 264906, 264907, 264908, 264909, 264511, 264512, 264910, 264591, 264592, 60432229, 264593, 264594, 33657402, 60433356, 264757, 60433438, 264596, 264758, 52646317, 21906754, 52644296, 265010, 264600, 264602, 264603, 264605, 264761, 264762, 264681, 264448, 264764, 264765, 264288, 264766, 264686, 264768, 264687, 264769, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 60170815, 264691, 33657023, 264692, 264693, 65274620, 27486284, 18108370, 264628, 264629, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264637, 264638, 52644332, 264639, 264558, 83373044, 60432113, 264564, 264565, 264566, 264486, 264567

2638	94326733 (5275, 5276)	Novel Protein sim. GBank gi4929889 gb AAD34105.1 AF151866 - (AF151866) CGI-110 protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	60424179, 52644507, 52646842, 18108398, 56182575, 22278995, 22278996, 35696286, 22278997, 22278999, 264259, 60432049, 29331822, 60424269, 29331826, 35696052, 29146498, 264905, 52644045, 56182435, 60433356, 33657402, 55812038, 55811386, 265019, 264288, 264769, 52644229, 56181562, 29148627, 29148629, 55811957, 29148784, 35695917, 265021, 52644150, 33657023, 65274620, 33657109, 35695763, 18108374, 55810764, 35696423, 55811576, 35695855, 60431850, 56182323, 60432113, 264404
2639	95361348 (5277, 5278)	Novel Protein sim. GBank gi2190007 dbj BAA203551 - (AB004109) phosphatidylserine synthase II [Oricetulus griseus]		synthase	264488, 29331825, 35696052, 264508, 264509, 264909, 264512, 33657402, 60433438, 264758, 85658542, 264600, 285020, 265021, 33657109, 264628, 35696423, 264555, 264639, 264563, 264564, 264565, 264566, 264488
2640	87781330 (5279, 5280)	Novel Protein sim. GBank gi3158516 (AF067617) - contains similarity to chromo (chromatin organization modifier) domains (Pfam: chromo.hmm, score: 17.76 and 27.94) and to helicases conserved C-terminal domain (Pfam: helicase_C.hmm, score: 67.00) [Caenorhabditis elegans]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	29331822, 29331826, 264908, 33109954, 265017, 265019, 21906788, 35695763, 264636, 264637, 18108387
2641	11669834 (5281, 5282)			UNCLASSIFIED	264828
2642	87412575 (5283, 5284)	Novel Protein sim. GBank gi2564955 (AF030001) - unknown [Mus musculus]		UNCLASSIFIED	264259, 29331822, 33657402, 265019, 264369, 264691, 264634, 56526486, 22279002
2643	87843981 (5285, 5286)	Novel Protein sim. GBank gi4490304 emb CAB38795.1  - (AL035678) putative protein [Arabidopsis thaliana]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	22278997, 264259, 29148499, 56182435, 264910, 265010, 18108351, 264682, 264683, 264369, 264684, 264685, 264686, 29148827, 264690, 33657109, 18108370, 263973, 18108374, 264634, 264557, 264558, 18108385, 264482
2644	88177671 (5287, 5288)	Novel Protein sim. GBank gi3789797 gb AAC67502.1  - (AF059569) actin binding protein MAYVEN [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	nucl_rept	264107, 264687
2645	17277228 (5289, 5290)			UNCLASSIFIED	265007
2646	94148542 (5291, 5292)	Novel Protein sim. GBank gi1708722 sp P49749 EVX2_MOUSE - HOMEBOX EVEN SKIPPED HOMOLOG PROTEIN 2 (EVX-2)		UNCLASSIFIED	264909, 264687, 264632, 83373044

2647	91212878 (5283, 5284)			UNCLASSIFIED	56182575, 22278996, 35696286, 22278998, 264259, 29331822, 56182181, 29331825, 60424269, 60432289, 35696052, 66712502, 264908, 265007, 55812038, 33109954, 21906754, 33657084, 265019, 264448, 264288, 56181562, 21906765, 21906766, 21906768, 21908769, 35695917, 265020, 265021, 52644150, 264693, 33657109, 33657349, 60431528, 18108374, 55810764, 35696423, 56182323, 60432113, 22279002, 264564
2648	87600587 (5295, 5296)				29146498, 56182435, 33109954, 265011, 264682, 55811957, 35695917, 264690, 263976, 18108377, 35696423, 60432113
2649	94128783 (5297, 5298)	Novel Protein sim. GBank gjl3041852 (AC004539) - unknown function; similar to Y09105 (PID:g1666171) [Homo sapiens]		UNCLASSIFIED	56182575, 35696286, 22278998, 29331824, 29331826, 60432289, 66712502, 56182435, 60170831, 60432228, 33657402, 33109954, 21906754, 265017, 264688, 21908765, 21906768, 60170815, 264693, 263967, 18108370, 263976, 60170394, 60432113, 22279002, 264563, 264685
2650	87297533 (5299, 5300)	Novel Protein sim. GBank gjl5360271 [dbj BAA81808.1] - (AB029335) HPET-3 [Halocynthia roretzi]			
2651	88088745 (5301, 5302)	Novel Protein sim. GBank gjl4240225 [dbj BAA74891.1] - (AB020675) KIAA0888 protein [Homo sapiens]	Contains protein domain (PF000054) - Laminin G domain	synthase	29331824, 29331826, 29331827, 265007, 55812038, 21906754, 18108366, 18108384, 22279002, 264567
2652	10343125 (5303, 5304)	Novel Protein sim. GBank gjl4493956 [emb CAB11123.2] - (Z98551) predicted using hexExon; MAL3P6.28 (PFC0846c), Hypothetical protein, len: 167 aa; Similarity to model organism hypothetical proteins (C.elegans, D.melanogaster, S.cerevisiae & S.pombe). C.elegans protein ZK287.5 (TR:...		UNCLASSIFIED	264692
2653	87798735 (5305, 5306)			UNCLASSIFIED	265018, 18108370, 18108387, 264566
2654	85103240 (5307, 5308)				60424179, 65274572, 56182575, 264259, 56182181, 264908, 56182435, 55811957, 35695917, 265021, 263976, 55810764, 65274791, 56182323, 83373044, 65274727
2655	91228018 (5309, 5310)	Novel Protein sim. GBank gjl3875272 [emb CAB02861] - (Z81051) predicted using GeneFINDER; similar to Zinc finger, C3HC4 type (RING finger); cDNA EST yk443h5.3 comes from this gene; cDNA EST yk443h5.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF000097) - Zinc finger, C3HC4 type (RING finger)	transcriptfactor	56182575, 56181686, 264092, 264259, 56182181, 60432289, 264907, 33657402, 55812038, 21906754, 87168559, 265017, 264448, 264369, 264288, 21906765, 21906766, 21906767, 21906768, 33657109, 18108370, 264628, 55811576, 264556, 264639, 83373044, 56526486, 264404, 60432113
2656	84562601 (5311, 5312)	Novel Protein sim. GBank gjl3043718 [dbj BAA25523] - (AB011169) KIAA0597 protein [Homo sapiens]			264693

2657	52561728 (5313, 5314)	Novel Protein sim. GBank gij5689509[djib BAA83038.1] - (AB028009) KIAA1086 protein [Homo sapiens]		dna_ma_bind	264693
2658	88062454 (5315, 5316)	Novel Protein sim. GBank gij3688089 (AC005757) - R32611_1 [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat		35686286, 264259, 29331822, 29331824, 29331826, 29331828, 265019, 264683, 21906768, 35695917, 264693, 35695855, 284637, 87168518, 264486, 264587
2659	87600755 (5317, 5318)	Novel Protein sim. GBank gij5420387[emb CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF01426) - BAH domain	UNCLASSIFIED	264909, 264910, 265018, 264369, 264769, 21906769, 264683, 263972, 18108388
2660	91718472 (5319, 5320)	Novel Protein sim. GBank gij728837[sp P39194 ALU7_HUMAN - III ALU SUBFAMILY SQ WARNING ENTRY IIII	Contains protein domain (PF00036) - EF hand	kinase	264488, 65274572, 35698286, 22278998, 22278999, 264259, 29331822, 29331824, 60432289, 29331826, 35696052, 264908, 56182435, 265008, 265009, 60433358, 264594, 265010, 265018, 55811150, 18108351, 264682, 264684, 264369, 264288, 264687, 21906765, 29148784, 35695917, 60170615, 52644150, 33657023, 33657109, 35696423, 35695855, 264556, 60170394, 18108385, 22278000, 22279002
2661	95342817 (5321, 5322)	Novel Protein sim. GBank gij4758048[ref NP_004739.1 pCPR8 - cell cycle progression 8 protein		glycoprotein	60432049, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 264906, 264909, 264593, 33109954, 265010, 265017, 265018, 265019, 264760, 264448, 264369, 264288, 21906765, 21906768, 265022, 264691, 33657023, 27486262, 60431528, 18108374, 35695855, 18108388, 264482, 264555, 264556, 264558, 264486
2662	80228739 (5323, 5324)	Novel Protein sim. GBank gij3674714[emb CAA91263] - (Z66494) similar to choline dehydrogenase; cDNA EST yk346d5.5 comes from this gene; cDNA EST yk346d5.3 comes from this gene [Caenorhabditis elegans]		dehydrogenase	264906, 264909, 264757, 264758, 264767, 264691, 33657023, 264638
2663	87780623 (5325, 5326)	Novel Protein sim. GBank gij1389670 (U58977) - Notch homolog Scalloped wings [Lucilia cuprina]	Contains protein domain (PF00008) - EGF-like domain	oncogene	35698286, 264509, 264595, 264288, 264685, 264686
2664	85518329 (5327, 5328)	Novel Protein sim. GBank gij4884408[emb CAB43311.1] - (AL050190) hypothetical protein [Homo sapiens]		UNCLASSIFIED	35698286, 22278999, 29331822, 35696052, 264906, 264907, 264909, 264510, 264511, 264512, 264593, 60433438, 265019, 264681, 21906765, 21906766, 21906767, 21906768, 265020, 265022, 35696423, 35695855, 22279002, 264482, 264486
2665	87770662 (5329, 5330)	Novel Protein sim. GBank gij5106956[gb AAD39906.1 AF11361] - (AF113615) FH1/FH2 domain-containing protein FHO5 [Homo sapiens]		UNCLASSIFIED	29331825, 265007, 264910, 60432229, 265019, 264288, 21906767, 264558, 22279002
2666	87626472 (5331, 5332)	Novel Protein sim. GBank gij5106956[gb AAD39906.1 AF11361] - (AF113615) FH1/FH2 domain-containing protein FHO5 [Homo sapiens]	Contains protein domain (PF01138) - 3' exoribonuclease family	nuclease	264907, 29331830, 264681, 264683, 264288, 35695855, 264632, 264556, 264557, 264558, 264559, 264563, 264565, 264567
2667	87422720 (5333, 5334)	Novel Protein sim. GBank gij2500570[sp Q17533 RNPH_CAEEL - RIBONUCLEASE PH-LIKE PROTEIN B0564.1			

2668	91216716 (5335, 5336)	Novel Protein sim. GBank gi15454186 ref NP_006327.1 pZYG - ZYG homolog		UNCLASSIFIED	56181686, 35696286, 22278998, 22278999, 56182181, 29331824, 60424269, 29331825, 35696052, 29331828, 66712502, 56182435, 60433356, 264758, 21906754, 55811386, 265011, 87168559, 265017, 265019, 55811150, 264448, 264369, 264288, 21906765, 21906766, 21906768, 55811957, 35695917, 265020, 265021, 33657023, 264692, 33657109, 35695763, 60431528, 18108374, 35696423, 55811576, 284634, 60431850, 83373044, 18108385, 87168518, 22279000, 264583, 264564
2669	95415721 (5337, 5338)	Novel Protein sim. GBank gi2147012 pir JC4899 - proline rich protein - rat			264489, 264689, 21906767, 65274572, 56182575, 21906768, 29148627, 21906769, 29148629, 35696286, 35695917, 22278998, 22278999, 265021, 265022, 80170615, 52644150, 60432049, 264259, 264691, 33657023, 264692, 29331822, 29331824, 29331825, 60432289, 33657109, 29331826, 29331827, 35696052, 29331828, 29146498, 29146499, 264905, 264908, 52644045, 264909, 56182435, 35696423, 65274791, 35695855, 265008, 264910, 264635, 60432229, 264592, 264638, 56182323, 60433356, 60170394, 60433436, 264559, 264595, 55812038, 33109954, 87168559, 60432113, 265019, 264448, 264369, 264684, 264288
2670	87613234 (5339, 5340)	Novel Protein sim. GBank gi1723523 sp Q10362 YD8B_SCHPO - HYPOTHETICAL 94.9 KD PROTEIN C22E12.11C IN CHROMOSOME I	Contains protein domain (PF00628) - PHD-finger		18108370, 263974
2671	91214936 (5341, 5342)	Novel Protein sim. GBank gi4788277 gb AAD29444.1 AF06425 - (AF064255) very long-chain acyl-CoA synthetase homolog 2; VLCS-H2 [Homo sapiens]		transport	52646842, 56994075, 264259, 29331822, 29331824, 29331825, 29331827, 33656970, 264509, 265008, 33109954, 21906754, 264682, 264288, 265021, 33657023, 33657109, 33657182, 27486261, 27486262, 27486285, 18108376, 18108385
2672	87399123 (5343, 5344)	Novel Protein sim. GBank gi4966346 gb AAD34677.1 AC00834 - (AC008341) Contains two PF01344 Kelch motif domains. [Arabidopsis thaliana]	Contains protein domain (PF01344) - Kelch motif	UNCLASSIFIED	264767

2873	87430749 (5345, 5346)	Novel Protein sim. GBank gi 5457337 emb CAB41505.2  - (A1236876) poly(ADP-ribose) polymerase-2 [Homo sapiens]	Contains protein domain (PF00644) - Poly(ADP-ribose) polymerase catalytic region.	polymerase	22278995, 22278996, 22278997, 22278999, 29331822, 29331824, 29331828, 35696052, 60433438, 87168474, 87168559, 265017, 265018, 265019, 264448, 21908768, 21908769, 265020, 265021, 33657109, 27486262, 35695763, 60431850, 60170394, 87168518, 264563
2874	94847721 (5347, 5348)	Novel Protein sim. GBank gi 4758824 ref NP_004280.1 pNRF3 - nuclear factor (erythroid-derived 2)-like 3	Contains protein domain (PF00170) - bZIP transcription factor	transcriptfactor	264488, 22278996, 35696286, 264091, 264259, 29331824, 29331826, 35696052, 264511, 55812038, 85658542, 264766, 21906765, 35695917, 264629, 35698423, 18108383, 87168518
2875	79563835 (5349, 5350)			UNCLASSIFIED	264691
2876	79628393 (5351, 5352)			UNCLASSIFIED	264906, 265008
2877	94328800 (5353, 5354)	Novel Protein sim. GBank gi 1079042 pir J52154 - acetyl-CoA synthetase - fruit fly (Drosophila melanogaster)	Contains protein domain (PF00501) - AMP-binding enzyme	synthase	18108394, 65274572, 56182575, 22278994, 22278995, 56994075, 22278996, 35696286, 22278997, 22278999, 264490, 264259, 52645080, 29331824, 29331825, 60432289, 29331827, 29331828, 35696052, 28146499, 29331830, 284908, 52644045, 265006, 265007, 265008, 265009, 60432229, 60433356, 60433438, 55812038, 265010, 265011, 87168559, 265017, 265019, 18108351, 264682, 264448, 264683, 264288, 21908765, 21906766, 21908767, 21906768, 21906769, 55811957, 265020, 265021, 60170615, 52644150, 264691, 33657023, 263967, 33657109, 27486284, 27486285, 33657348, 35695763, 18108370, 18108374, 18108377, 55811576, 35698423, 35695855, 83373044, 18108387, 22279000, 22278002, 264564



2678	95001604 (5355, 5356)	Novel Protein sim. GBank g 186760 p j A40465 - alanine transaminase (EC 2.6.1.2), cytosolic - human		UNCLASSIFIED	264488, 263994, 264489, 18108394, 52646842, 35696286, 22278998, 264259, 29331825, 35698052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 264512, 265007, 265008, 264910, 265009, 264593, 264594, 264596, 264758, 33109954, 52646317, 21906754, 265010, 265011, 87168559, 264600, 264601, 264602, 264603, 265017, 265019, 264605, 264760, 264762, 18108351, 264763, 264682, 264683, 264764, 264288, 264766, 264687, 264768, 264769, 264689, 21906765, 21906767, 35695917, 265020, 265021, 52644150, 264691, 33657023, 264693, 18108364, 18108365, 33657109, 33857349, 264628, 18108374, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 18108385, 264563, 264584, 264565, 264566, 264486, 264567
2679	95361544 (5357, 5358)	Novel Protein sim. GBank g 1709233 p P07514 NC5R_BOVIN - NADH-CYTOCHROME B5 REDUCTASE	Contains protein domain (PF00970) - FAD/NAD-binding Cytochrome reductase	reductase	264488, 22278998, 35696286, 264259, 29331826, 29331827, 29331828, 264809, 56182435, 264113, 264511, 265008, 60433438, 264758, 85658542, 87168474, 265011, 265017, 265018, 264288, 21906766, 21906767, 21906768, 21906769, 55811957, 265021, 264690, 33657023, 55810764, 35696423, 55811576, 264631, 18108381, 60170394, 83373044, 87168518, 264566
2680	87800356 (5359, 5360)	Novel Protein sim. GBank g 4589604 db BAA76824.1  - (AB023197) KIAA0980 protein [Homo sapiens]	Contains protein domain (PF00036) - EF hand	strut	264259, 264102, 264905, 264908, 265007, 265008, 33109954, 265011, 18108351, 264768, 33657023, 20281149, 263972, 264630, 264635, 264638
2681	90933844 (5361, 5362)	Novel Protein sim. GBank g 1728837 p P39194 ALU7_HUMAN - !III ALU SUBFAMILY SQ WARNING ENTRY IIII		kinase	264489, 22278995, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265006, 264512, 264910, 264594, 60433438, 264758, 264603, 264604, 265019, 264605, 264760, 264764, 264687, 264768, 264769, 21906769, 35695917, 265020, 33657023, 264631, 264635, 264637, 264638, 264639, 264486
2682	94138934 (5363, 5364)	Novel Protein sim. GBank g 423468 p j JQ1974 - HTF9-C protein - mouse	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	35696286, 264908, 55811388, 265017, 55811150, 55811957, 35695917, 60431528, 55810764, 55811576, 35696423, 65274791, 55526486
2683	87774405 (5365, 5366)	Novel Protein sim. GBank g 5114351 g JAD40286.1  - (AF156271) RING finger protein terf [Homo sapiens]	Contains protein domain (PF00622) - SPRY domain	interleukinrecept	264909, 264769, 264635, 264636

2684	85787151 (5367, 5368)	Novel Protein sim. GBank gi 4886469 emb CAB43385.1  - (AL050284) hypothetical protein [Homo sapiens]			264593	
2685	88054289 (5369, 5370)	Novel Protein sim. GBank gi 3342729 (AC005331) - R31341.2 [Homo sapiens]		UNCLASSIFIED		
2686	87628690 (5371, 5372)	Novel Protein sim. GBank gi 4650844 dbj BAA77027.1  - (AB026190) Kelch motif containing protein [Homo sapiens]	Contains protein domain (PF01344) - struct Kelch motif		264259, 29331822, 60432289, 29331827, 29331830, 264909, 264512, 264596, 264769, 264534, 264555, 264556, 264557, 264558, 60170394, 264559, 264488	
2687	87998183 (5373, 5374)	Novel Protein sim. GBank gi 5281314 gb AAD41475.1 AF133123 - (AF133123) transcription factor IIC102 [Homo sapiens]	Contains protein domain (PF00515) - transcriptfactor TPR Domain		18108394, 18108396, 22278996, 35696286, 22278997, 29331826, 29331828, 66712502, 21906754, 265011, 264760, 264761, 264763, 264889, 21906765, 35696423, 264559, 18108385, 264563	
2688	78959584 (5375, 5376)	Novel Protein sim. GBank gi 3880023 emb CAA97339  - (Z73098) Similarity to yeast hypothetical protein (Swiss Prot accession number Q09695); cDNA EST EMBL:D72982 comes from this gene; cDNA EST EMBL:D75947 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00560) - Leucine Rich Repeat		264908, 264760	
2689	84122440 (5377, 5378)	Novel Protein sim. GBank gi 3880023 emb CAA97339  - (Z73098) Similarity to yeast hypothetical protein (Swiss Prot accession number Q09695); cDNA EST EMBL:D72982 comes from this gene; cDNA EST EMBL:D75947 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00560) - Leucine Rich Repeat		22278997, 22278998, 22278999, 29331824, 35696052, 264908, 264908, 58182435, 264512, 264910, 265009, 60433438, 21906754, 18108351, 264682, 264683, 264787, 21906765, 21906766, 21906768, 33657023, 33657182, 27486262, 27486264, 33657349, 18108370, 18108372, 18108374, 35696423, 35695855, 18108385, 22278002	
2690	88003055 (5378, 5380)	Novel Protein sim. GBank gi 2477513 (AC002398) - F25965.3 [Homo sapiens]	Contains protein domain (PF00620) - RhoGAP domain	struct		
2691	81219241 (5381, 5382)	Novel Protein sim. GBank gi 4107276 emb CAA87130  - (X98506) acetyl-CoA synthetase [Solanum tuberosum]		synthase	65274572, 56182575, 264259, 29331822, 29331824, 29331826, 29331828, 264112, 265009, 55812038, 264596, 33109954, 265017, 264448, 264288, 264369, 264684, 21906769, 60170615, 60431528, 55810764, 264634, 264636, 264556, 264637, 22278002, 264584, 264566	
2692	84111914 (5383, 5384)	Novel Protein sim. GBank gi 3513303 (AC005594) - R26984.1 [Homo sapiens]	Contains protein domain (PF00326) - Proyl oligopeptidase family	peptidase		
2693	20438807 (5385, 5386)	Novel Protein sim. GBank gi 3122400 sp Q35682 MUG_MOUSE - MYELOID UPREGULATED PROTEIN		UNCLASSIFIED	264592	
2694	84111916 (5387, 5388)	Novel Protein sim. GBank gi 3122400 sp Q35682 MUG_MOUSE - MYELOID UPREGULATED PROTEIN			264559	
2695	95345513 (5389, 5390)	Novel Protein sim. GBank gi 4972740 gb AAD34785.1  - (AF132177) unknown [Drosophila melanogaster]		collagen	35696286, 56994075, 22278999, 264259, 35696052, 29331830, 265011, 264288, 56181562, 264690, 264692, 33657023, 27486262, 263976, 18108376, 35696423, 35695855, 60170394, 83373044, 56526486, 22279000, 22279002, 264566	
2696	87874040 (5391, 5392)	Novel Protein sim. GBank gi 728831 sp P39188 ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII		synthase	264594, 21906768, 18108370, 18108372	

2697	91638472 (5393, 5394)	Novel Protein sim. GBank gij5689473[dbj BAA83020.1] - (AB028991) KIAA1068 protein [Homo sapiens]		UNCLASSIFIED	35696286, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 265006, 265007, 264512, 265009, 60170831, 60433356, 264595, 264758, 87168474, 265010, 265011, 87168559, 264601, 265017, 265018, 265019, 264761, 18108351, 264448, 264288, 264766, 264688, 264689, 21906765, 21906768, 265020, 265021, 60170615, 33657109, 18108376, 35696423, 35695855, 264555, 264558, 60170394, 264559, 18108387, 56526486, 87168518, 22279002, 264563, 264482
2698	94325891 (5395, 5396)	Novel Protein sim. GBank gij841318 (U22818) - mutant sterol regulatory element binding protein-2 [Cricetulus griseus]	Contains protein domain (PF00412) - LIM domain containing proteins	transcriptfactor	22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 33656970, 264806, 28331830, 264909, 52644045, 264910, 60433356, 33657402, 33109954, 265017, 265018, 265019, 264288, 21906765, 21906766, 21906767, 21906769, 29148629, 35695917, 265021, 265022, 52644150, 33657023, 33657182, 27486261, 35696423, 65274791, 264638, 60432113, 22279000
2699	87780650 (5397, 5398)			UNCLASSIFIED	264768, 18108357, 264690, 264691
2700	94139836 (5399, 5400)	Novel Protein sim. GBank gij5174395[ref NP_006006.1 pB120 - Brain protein 120		UNCLASSIFIED	60424269, 56182435, 60432229, 60433438, 55811386, 265017, 55811150, 264448, 56181582, 55811957, 264693, 33657109, 60431528, 264629, 55810764, 55811576, 65274791, 60431850, 60432113
2701	94148584 (5401, 5402)	Novel Protein sim. GBank gij1572801 (U70854) - F38A5.1 gene product [Caenorhabditis elegans]			18108394, 52845156, 35696286, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 29146489, 265006, 60433356, 33657402, 60433438, 264595, 33657084, 18108351, 264288, 264769, 18108359, 21906768, 35695917, 33657023, 27486261, 18108374, 18108379, 35696423, 18108382, 83373044, 18108384, 18108388, 60432113, 22279000
2702	57295368 (5403, 5404)	Novel Protein sim. GBank gij2605967 (AF030027) - 24 [Equine herpesvirus 4]		UNCLASSIFIED	29331828, 264512, 264555, 264556, 264557, 264558, 264559
2703	87649514 (5405, 5406)	Novel Protein sim. GBank gij5689399[dbj BAA82983.1] - (AB028954) KIAA1031 protein [Homo sapiens]	Contains protein domain (PF00642) - Zinc finger C-x8-C-x5-C-x3-H type (and similar).		60432289, 265007, 21906765, 21906768, 265021, 264563

2704	87649515 (5407, 5409)	Novel Protein sim. GBank gi 4335694 gb AAB63294  - (AF008554) Implantation-associated protein [Rattus norvegicus]			264488, 22278995, 22278998, 29331828, 29146499, 264905, 264908, 264907, 52644045, 264511, 33657402, 264600, 264602, 265017, 264605, 264761, 18108351, 264764, 264687, 264769, 265021, 264691, 264692, 18108362, 264693, 18108370, 18108374, 264634, 264635
2705	87771745 (5409, 5410)				264489, 264509, 264511, 264512, 264910, 264593, 87168474, 264604, 264288, 264687, 264769, 264638, 264566, 264486
2706	94326789 (5411, 5412)	Novel Protein sim. GBank gi 3255952 emb CAA16821.1  - (AL021728) /prediction=(method: /match=(desc: [Drosophila melanogaster])	UNCLASSIFIED		264488, 52646842, 65274572, 22278994, 56994075, 22278997, 264259, 29331824, 29331825, 29331826, 29331828, 33656970, 264907, 264908, 264909, 52644045, 58182435, 265006, 265007, 60433438, 55812038, 21908754, 52644296, 265010, 264601, 265017, 265019, 264681, 264448, 264682, 264288, 264686, 264687, 264688, 21906766, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 264690, 264691, 33657023, 264692, 264693, 65274620, 27486284, 263972, 18108374, 18108377, 264635, 264636, 264556, 60170394
2707	88089838 (5413, 5414)	Novel Protein sim. GBank gi 3417294 (AC004381) - Unknown gene product [Homo sapiens]			83373044, 65274727, 87168518, 22279000, 22278996, 22278998, 56182435, 21906754, 87168559, 265017, 264448, 52645129
2708	91011351 (5415, 5416)	Novel Protein sim. GBank gi 545790 bbs 147178 - DARPP-32=dopamine and cAMP-regulated phosphoprotein [human, brain, Peptide, 204 aa]	UNCLASSIFIED		65274572, 264259, 29331822, 29331825, 60432289, 29331826, 29331827, 29331828, 264909, 264510, 265007, 264910, 60433356, 60433438, 33109854, 265010, 265011, 264389, 264288, 264765, 264693, 264565
2709	94853888 (5417, 5418)	Novel Protein sim. GBank gi 3169705 (AC004780) - F17127_1 [Homo sapiens]	UNCLASSIFIED		29331822, 18108370, 18108374, 83373044
2710	87627878 (5419, 5420)	Novel Protein sim. GBank gi 4468311 emb CAB37892  - (AL031432) dJ465N24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]			29331824, 264759, 264693, 18108382, 18108388

2711	94111920 (5421, 5422)	Novel Protein sim. GBank gij3122400[sp]O35682[MG_MOUSE - MYELOID UPREGULATED PROTEIN		UNCLASSIFIED	264488, 264687, 52645156, 264769, 21908764, 21906765, 21906767, 21906768, 21906769, 55811957, 56994075, 22278997, 22278998, 265020, 265021, 264690, 264259, 264691, 264692, 33657023, 29331822, 29331824, 60424269, 29331826, 33657182, 29331827, 27486262, 33657349, 264508, 264905, 264907, 60431528, 264908, 264909, 55810764, 35696423, 65274791, 35695855, 265007, 264910, 60431850, 60432229, 264557, 264558, 55812038, 33108954, 18108385, 21906754, 33657084, 87168518, 87168474, 87168559, 60432113, 265017, 22279000, 265018, 265019, 22279002, 264760, 55811150, 264681, 18108351, 264565, 264764, 264566, 264288, 264766 264488, 35696286, 22278998, 264259, 29331824, 60432289, 35696052, 264508, 264906, 66712502, 52644045, 265006, 60432229, 33657402, 60433358, 265010, 265019, 18108351, 264681, 264288, 264685, 21906765, 21906766, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 264691, 264692, 33657023, 33657109, 33657182, 33657349, 18108370, 18108374, 35696423, 35695855, 264555, 52644332, 56182323, 87168518, 60432113 66714117, 264906, 264563
2712	94312071 (5423, 5424)	Novel Protein sim. GBank gij5081315[gb]AAD39343.1[AF07660] - prediabetic NOD sera-reactive autoantigen [Mus musculus]	Contains protein domain (PF00515) - TPR Domain	transferase	
2713	88003064 (5425, 5426)	Novel Protein sim. GBank F25965_3 [Homo sapiens]		UNCLASSIFIED	
2714	13528218 (5427, 5428)			UNCLASSIFIED	264636
2715	94122454 (5429, 5430)	Novel Protein sim. GBank gij4321968[gb]AAD15897] - (AF067430) Smarce1-related protein [Mus musculus]		UNCLASSIFIED	264508, 264905, 264907, 264908, 264909, 264910, 264758, 265011, 264760, 264764, 264288, 264766, 264768, 264636
2716	88003068 (5431, 5432)	Novel Protein sim. GBank gij2477513 (AC002398) - F25965_3 [Homo sapiens]		glycoprotein	264091, 264259, 29331822, 66714117, 264908, 264369, 264693, 264556, 264563
2717	80077461 (5433, 5434)	Novel Protein sim. GBank gij3327046[dbj]BAA31591] - (AB014516) KIAA0616 protein [Homo sapiens]			264593, 264558
2718	79604062 (5435, 5436)				264693
2719	88180423 (5437, 5438)	Novel Protein sim. GBank gij746495 (U233515) - weakly similar to gastrula zinc finger protein [Caenorhabditis elegans]		UNCLASSIFIED	29331822, 87168559, 265019, 265021, 52644150, 264691

2720	95086242 (5438, 5440)	Novel Protein sim. GBank gi 1335873 (U46690) - ATP-dependent RNA helicase [Mus musculus]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	18108374, 60424179, 264489, 56182435, 21908765, 21908766, 35698423, 22278997, 265020, 265022, 265008, 265008, 264092, 264636, 60432229, 264691, 264692, 33657023, 264693, 33657402, 83373044, 29331824, 18108366, 60424269, 29331826, 18108385, 52645129, 21906754, 35696052, 29331828, 87168474, 264100, 265010, 265011, 265019, 22279002, 264905, 264482, 264563, 264906, 18108351, 264681, 18108370, 29331830, 264908, 66712502, 52644045, 264809, 264828, 18108354, 22278995, 35696286, 264259, 29331822, 29331824, 66714117, 29331826, 264906, 60433438, 265017, 18108351, 264448, 264288, 264769, 21906766, 265021, 33657109, 263969, 60431528, 264629, 55811576, 65274791, 35695855, 264831, 264837, 60170394, 56182323, 22279000, 35696286, 22278999, 21906754, 265017, 264762, 264288, 21906765, 21906767, 21906768, 35695917, 18108362, 27486262, 35695855, 264558, 264559
2721	95345523 (5441, 5442)	Novel Protein sim. GBank gi 4928663 gb AAD34092.1 AF15185 - (AF151855) CGI-97 protein [Homo sapiens]	Contains protein domain (PF011172) - Uncharacterized protein family UPF0023	18108374, 60424179, 264489, 56182435, 21908765, 21908766, 35698423, 22278997, 265020, 265022, 265008, 265008, 264092, 264636, 60432229, 264691, 264692, 33657023, 264693, 33657402, 83373044, 29331824, 18108366, 60424269, 29331826, 18108385, 52645129, 21906754, 35696052, 29331828, 87168474, 264100, 265010, 265011, 265019, 22279002, 264905, 264482, 264563, 264906, 18108351, 264681, 18108370, 29331830, 264908, 66712502, 52644045, 264809, 264828, 18108354, 22278995, 35696286, 264259, 29331822, 29331824, 66714117, 29331826, 264906, 60433438, 265017, 18108351, 264448, 264288, 264769, 21906766, 265021, 33657109, 263969, 60431528, 264629, 55811576, 65274791, 35695855, 264831, 264837, 60170394, 56182323, 22279000, 35696286, 22278999, 21906754, 265017, 264762, 264288, 21906765, 21906767, 21906768, 35695917, 18108362, 27486262, 35695855, 264558, 264559
2722	91638807 (5443, 5444)	Novel Protein sim. GBank gi 3212997 gb AAC23434.1  - (AC004997) match to ESTs AA667999 (NID:g2626700), AA165465 (NID:g1741481), Z45871 (NID:g575105), and T84026 (NID:g712314); similar to various tre-like proteins including: AF040654 (PID:g2746883), D13844 (PID:g2104571), AL02114...	Contains protein domain (PF00566) - TBC domain	18108374, 60424179, 264489, 56182435, 21908765, 21908766, 35698423, 22278997, 265020, 265022, 265008, 265008, 264092, 264636, 60432229, 264691, 264692, 33657023, 264693, 33657402, 83373044, 29331824, 18108366, 60424269, 29331826, 18108385, 52645129, 21906754, 35696052, 29331828, 87168474, 264100, 265010, 265011, 265019, 22279002, 264905, 264482, 264563, 264906, 18108351, 264681, 18108370, 29331830, 264908, 66712502, 52644045, 264809, 264828, 18108354, 22278995, 35696286, 264259, 29331822, 29331824, 66714117, 29331826, 264906, 60433438, 265017, 18108351, 264448, 264288, 264769, 21906766, 265021, 33657109, 263969, 60431528, 264629, 55811576, 65274791, 35695855, 264831, 264837, 60170394, 56182323, 22279000, 35696286, 22278999, 21906754, 265017, 264762, 264288, 21906765, 21906767, 21906768, 35695917, 18108362, 27486262, 35695855, 264558, 264559
2723	87387732 (5445, 5446)		UNCLASSIFIED	264508, 264509, 264906, 264908, 264910, 55812038, 264766, 264687, 264628, 264636, 264486
2724	87638563 (5447, 5448)	Novel Protein sim. GBank gi 4580881 gb AAD27730.1 AF13295 - (AF132955) CGI-21 protein [Homo sapiens]	ubiquitin	18108396, 22278999, 20281099, 29331824, 29331826, 60432289, 29331828, 60170831, 60432229, 60433438, 18108351, 264682, 21906766, 21906767, 21906769, 35695917, 33657023, 33657109, 18108372, 18108374, 35695855, 22279000, 22279002
2725	94853991 (5449, 5450)	Novel Protein sim. GBank gi 3169705 (AC004780) - F17127_1 [Homo sapiens]	UNCLASSIFIED	264488, 52644507, 264259, 29331827, 21906754, 285011, 18108351, 264448, 264288, 264685, 264689, 35695917, 265020, 33657182, 27486261, 18108370, 18108374, 35696423, 18108385, 22279000
2726	86880599 (5451, 5452)	Novel Protein sim. GBank gi 3342738 (AC005328) - R26860_1, partial CDS [Homo sapiens]	MHC	264488, 264828, 264685

2727	81010470 (5453, 5454)	Novel Protein sim. GBank gi 731267 sp P39219 RLUA_ECOLI - RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE A (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE)	Contains protein domain (PF00849) - RNA pseudouridylate synthase	- synthase	65274572, 56182575, 22278984, 56994075, 22278986, 22278997, 22278999, 60432049, 29331822, 29331824, 66714117, 29331826, 29331827, 35696052, 29331828, 33856970, 264508, 66712502, 264910, 33657402, 60433438, 264758, 55812038, 21906754, 33657084, 55811386, 265018, 265019, 264767, 21906765, 21906767, 21906769, 55811957, 35695917, 52644150, 33657023, 33657109, 33657182, 27486281, 27486282, 27486285, 33657349, 55811576, 35696423, 35695855, 264630, 60431850, 264636, 56182323, 87168518, 60432113, 22279000, 264564, 264565
2728	94126022 (5455, 5456)	Novel Protein sim. GBank gi 3880433 emb CAA91399  - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein; cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	18108394, 56182435, 21906767, 55811957, 35695855, 265021, 264690, 264556, 264259, 264557, 29331822, 264559, 264448, 264288
2729	94126024 (5457, 5458)	Novel Protein sim. GBank gi 2408095 emb CAB16300  - (Z99168) putative RNA splicing protein [Schizosaccharomyces pombe]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	65274572, 264259, 60432289, 66712502, 56182435, 264448, 264288, 264369, 55811957, 265021, 264557, 60432113
2730	94126026 (5459, 5460)	Novel Protein sim. GBank gi 3880433 emb CAA91399  - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein; cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	- transport	264887, 29331824, 29331826, 35696052, 264107, 56182435, 265008, 265009, 264592, 60431735, 265011, 264601, 265017, 18108351, 264288, 29148627, 55811957, 265021, 264690, 18108368, 18108374, 264557, 264558, 284559, 18108387, 56526486, 264566, 264486
2731	87723022 (5461, 5462)	Novel Protein sim. GBank gi 1723239 sp Q10169 YAU8_SCHPO - HYPOTHETICAL 35.7 KD PROTEIN C26A3.1 IN CHROMOSOME I	Contains protein domain (PF00795) - Carbon-nitrogen hydrolase		264259, 35696052, 265008, 264758, 264762, 264448, 264288, 29148627, 21908769, 87168518, 22279002
2732	94126028 (5463, 5464)	Novel Protein sim. GBank gi 3880433 emb CAA91399  - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein; cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	264887, 264489, 18108358, 56182435, 264689, 35696423, 55811957, 265021, 265006, 265008, 264910, 265009, 264690, 264555, 264259, 264556, 264557, 264558, 264559, 18108383, 33657109, 87168518, 265010, 264601, 60432113, 265017, 264905, 264448, 263972, 264369, 264567
2733	87383060 (5465, 5466)			UNCLASSIFIED	29331825, 264509, 264909
2734	94140286 (5467, 5468)	Novel Protein sim. GBank gi 4519621 dbj BAA75670.1  - (AB017614) OASIS protein [Mus musculus]	Contains protein domain (PF00170) - bZIP transcription factor	- transcription factor	60424179, 52644507, 56182575, 264259, 29331828, 264907, 264510, 264910, 60433356, 265019, 55811150, 264681, 264763, 264687, 33657182, 18108370, 60431528, 60431850, 56182323, 83373044

2735	87712336 (5469, 5470)	Novel Protein sim. GBank gij3850569 (AC005278) - ESTs gbj121278, gbj145403, and gbjAA586113 come from this gene. [Arabidopsis thaliana]		glycoprotein	22278996, 60432289, 29331827, 28146498, 264108, 284909, 264112, 33657402, 87168474, 285017, 264762, 264448, 264764, 264684, 21906765, 264693, 33657109, 263976, 284636, 264638, 264557, 22279000, 22279002, 264567
2736	80247655 (5471, 5472)			UNCLASSIFIED	264905, 264628, 264629, 263978, 264632, 264564
2737	87604526 (5473, 5474)				264488
2738	85731808 (5475, 5476)	Novel Protein sim. GBank gij2550501[dbj BA22896] - (D63850) hepatoma-derived growth factor. [Mus musculus]			264488, 265009, 264768, 264691
2739	94319834 (5477, 5478)	Novel Protein sim. GBank gij5420387[emb CA846679.1] - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264684, 83373044, 284588
2740	94148762 (5479, 5480)	Novel Protein sim. GBank gij3417386[emb CAA75495] - (Y15197) microtubule-associated protein, MAP-115 [Mus musculus]		UNCLASSIFIED	264488, 56182575, 22278995, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 29146498, 29331830, 265006, 265007, 265009, 60432229, 33657402, 55812038, 87168474, 265010, 265011, 265017, 265018, 265019, 264605, 264681, 264288, 264369, 52644229, 21906765, 21906768, 21906767, 21906768, 21906769, 265020, 265022, 264691, 264692, 33657109, 18108370, 18108374, 55810764, 35695855, 264634, 60431850, 264639, 56182323, 18108382, 18108385, 65274727, 22279002, 264584
2741	88047518 (5481, 5482)	Novel Protein sim. GBank gij3242764 (AC005154) - similar to protein U28928 (PID:g861306) [Homo sapiens]		UNCLASSIFIED	22278996, 52644045, 52644229, 21906768, 21906769, 265020, 60170615, 264691
2742	87648644 (5483, 5484)	Novel Protein sim. GBank gij4758412[ref NP_004472.1]pGALN - UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetyl-galactosaminyltransferase 2 (GalNAc-T2)	Contains protein domain (PF00852) - Similarity to lectin domain of ricin beta-chain, 3 copies.	transferase	264258, 264805, 264758, 55812038, 264369, 29148627
2743	87627991 (5485, 5486)	Novel Protein sim. GBank gij4468311[emb CA837992] - (AL031432) cJ465N24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]		UNCLASSIFIED	35696286, 264259, 264906, 264908, 265006, 60433438, 285017, 18108351, 264448, 264764, 264288, 21906765, 21906767, 264690, 264691, 264693, 263969, 263971, 35695855, 264637, 264558, 18108382, 60432113
2744	94126030 (5487, 5488)	Novel Protein sim. GBank gij3880433[emb CAA91399] - (Z66521) similar to mitochondrial RNA splicing MSR4 like protein: cDNA EST EMBL:CO9217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	kinase	18108374, 264488, 56182435, 21906765, 35696423, 35695817, 35695855, 265020, 295021, 264511, 265009, 264490, 264556, 264259, 264557, 56182323, 264558, 264559, 18108383, 29331824, 18108385, 33657109, 29331826, 21906754, 29331827, 29331828, 33657349, 87168518, 265018, 264905, 264482, 264448, 264486, 264369, 264288



2745	87740125 (5489, 5490)	Novel Protein sim. GBank gij4405795[gb AAD19826] - (AF038963) RNA helicase [Homo sapiens]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	35696286, 264509, 264905, 264907, 264908, 264909, 264510, 264512, 265008, 264758, 264601, 265017, 264604, 264763, 264288, 264686, 264769, 264693, 35696423, 35695855, 264634, 264636, 264563, 264564, 264565
2746	95418601 (5491, 5492)	Novel Protein sim. GBank gij4758738[ref NP_004680.1 pMTA1 - metastasis associated 1	Contains protein domain (PF00320) - GATA zinc finger	UNCLASSIFIED	22278996, 22278998, 22278999, 29331822, 29331826, 29331827, 35696052, 29331828, 264905, 264908, 264907, 264908, 264909, 52644045, 285006, 60170831, 284598, 55812038, 265018, 264683, 264288, 21906765, 21906767, 21906768, 21906769, 265020, 264690, 33657023, 264693, 33657109, 18108368, 18108374, 264558, 18108385, 22279000, 264563
2747	94112677 (5493, 5494)	Novel Protein sim. GBank gij4557803[ref NP_000262.1 pNPC1 - Niemann-Pick disease, type C1		glycoprotein	264569, 52644507, 18108394, 22278995, 35696286, 22278997, 22278999, 52645080, 29331824, 56182181, 29331826, 29331827, 35696052, 284907, 264908, 264909, 265009, 33109954, 55811386, 87168474, 265010, 87168559, 264603, 265019, 264760, 264686, 264768, 21906769, 35695917, 60170615, 264692, 33657023, 52645129, 27486264, 60431528, 18108374, 35696423, 35695855, 264556, 56182323, 18108385, 264482
2748	91214983 (5495, 5496)	Novel Protein sim. GBank gij4191272[emb CAA09984] - (AJ012295) apaG protein [Rhizobium etli]	Contains protein domain (PF00646) - F-box domain.		65274572, 29331828, 264112, 264511, 265019, 264760, 264767, 264768, 264769, 21906768, 21906769, 265020, 27486262, 5626486, 87168518, 22279000
2749	87346307 (5497, 5498)				264258, 264908, 264510, 265008, 265009, 264760, 264369, 264768, 264563
2750	87336344 (5499, 5500)	Novel Protein sim. GBank gij1872498 (U74297) - PIUS [Oryctolagus cuniculus]		UNCLASSIFIED	264488, 52644507, 18108396, 56994075, 264259, 29331825, 29331826, 29331827, 29331828, 264508, 265009, 264910, 264591, 264595, 33857084, 265011, 265019, 18108351, 264288, 264686, 264769, 264689, 55811957, 264693, 27486264, 18108370, 18108374, 264558, 18108385, 264482, 264563
2751	87057465 (5501, 5502)			UNCLASSIFIED	29331822, 29331824, 265017, 33657023
2752	88062675 (5503, 5504)	Novel Protein sim. GBank gij3041859 (AC004534) - OG-2 homeodomain protein-like; similar to U65087 (PID:g1575526) [Homo sapiens]	Contains protein domain (PF00046) - Homeobox domain	homeobox	

2753	94138972 (5505, 5506)	Novel Protein sim. GBank gjl3851648 (AF098301) - neural F-box protein NFB42 [Rattus norvegicus]	Contains protein domain (PF00646) - F-box domain.	UNCLASSIFIED	58182575, 56994075, 22278998, 29331822, 29331824, 29331825, 29331826, 265007, 264593, 55812038, 33109954, 18108351, 264288, 56181562, 21906767, 21908768, 265021, 264693, 18108374, 65274791, 264632, 56182323, 22279002, 264563, 264567
2754	94115513 (5507, 5508)	Novel Protein sim. GBank gjl535428 (U13736) - calmodulin-like protein [Pisum sativum]	Contains protein domain (PF00036) - EF hand	struct	22278999, 66714117, 29331827, 35695052, 29331828, 264908, 264908, 264909, 265011, 265017, 265018, 265019, 264288, 21906765, 21906767, 265022, 33657023, 264693, 56182323, 18108382, 22279000
2755	88001472 (5509, 5510)	Novel Protein sim. GBank gjl2898653 (AC004510) - R30385_2 [Homo sapiens]	Contains protein domain (PF00098) - Zinc finger, C2H2 type	transcriptfactor	
2756	11465908 (5511, 5512)	Novel Protein sim. GBank gjl173539 (U30473) - putative src-like adapter protein; non-catalytic src-like adapter protein containing SH3 and SH2 domains; homolog of mouse SLAP; Method: conceptual translation supplied by author [Homo sapiens]	Contains protein domain (PF00017) - Src homology domain 2	UNCLASSIFIED	264594
2757	95361590 (5513, 5514)	Novel Protein sim. GBank gjl3608372 (AF053768) - brain specific cortactin-binding protein CBP80 [Rattus norvegicus]		eph	85658542, 264693
2758	78637846 (5515, 5516)	Novel Protein sim. GBank gjl2072200 (U94863) - p40 [Borna disease virus]		struct	264693
2759	91005312 (5517, 5518)	Novel Protein sim. GBank gjl4814573(emb)CAB43885.1] - (AL050390) hypothetical protein [Homo sapiens]	Contains protein domain (PF00023) - Ankr repeat	UNCLASSIFIED	65274572, 35696286, 66714117, 29331828, 264508, 56182435, 21906754, 55811957, 264629, 264638, 56182323, 22279002, 264908
2760	79824798 (5518, 5520)	Novel Protein sim. GBank gjl3136150 (AF050182) - PERIOD 3 [Mus musculus]	Contains protein domain (PF00989) - PAS domain	nuclease	22278998, 29331822, 29331830, 265010, 265019, 264288, 21906765, 21906768, 21906769, 265020, 56182323, 22279002, 264563
2761	87639597 (5521, 5522)	Novel Protein sim. GBank gjl2905643 (AF045244) - ribitol kinase [Klebsiella pneumoniae]	Contains protein domain (PF01352) - KRAB box	struct	18108394, 22278998, 264906, 264909, 265006, 265007, 264757, 265010, 265011, 265017, 265019, 18108351, 264448, 264683, 264686, 264768, 265020, 265021, 265022, 264691, 18108362, 264693, 18108365, 33657109, 18108368, 18108370, 18108381, 18108382, 18108384, 18108386, 87168518, 264369, 35696423
2762	87592698 (5523, 5524)	Novel Protein sim. GBank gjl3136150 (AF050182) - PERIOD 3 [Mus musculus]	Contains protein domain (PF00370) - FGYY family of carbohydrate kinases	transcriptfactor	52646842, 264259, 29331822, 29331825, 29331826, 29331828, 33656970, 264905, 264907, 29331830, 265006, 265009, 21906754, 265019, 264448, 21906769, 27465262, 56182323, 56526486, 87168518, 264487
2763	87539868 (5525, 5526)	Novel Protein sim. GBank gjl2905643 (AF045244) - ribitol kinase [Klebsiella pneumoniae]			
2764	94305140 (5527, 5528)	Novel Protein sim. GBank gjl2905643 (AF045244) - ribitol kinase [Klebsiella pneumoniae]			

2765	94315105 (5528, 5530)	Novel Protein sim. GBank gi 4688672 emb CAA17688.2  - (AL022018) /prediction=(method:; /prediction=(method:; /match=(desc: [Drosophila melanogaster])			264488, 65274572, 22278995, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264509, 284905, 264908, 264907, 264908, 264909, 284512, 264910, 285009, 264592, 264595, 264758, 55812038, 33109954, 265010, 87168559, 264600, 285018, 264760, 264761, 264762, 264763, 264448, 264764, 264288, 264766, 264767, 264768, 56181562, 21906764, 21906765, 21906768, 35695917, 265021, 264691, 264692, 33657023, 33657109, 264628, 18108370, 264629, 18108374, 55811576, 35696423, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 264558, 264639, 83373044, 87168518, 22279000, 22279002, 264563, 264482
2766	94315109 (5531, 5532)	Novel Protein sim. GBank gi 5441611 emb CAB46854.1  - (AJ388555) hypothetical protein [Canis familiaris]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_ma_bind	264488, 65274572, 60432289, 264907, 264909, 264511, 264512, 60433356, 264288, 264685, 264689, 35695917, 265022, 264693, 264628, 65274791, 264635, 264555, 264556, 264557, 264638, 264558, 264559, 83373044, 60432113
2767	80204297 (5533, 5534)	Novel Protein sim. GBank gi 1079451 pir A55463 - tropomodulin, skeletal muscle - chicken		struct	264112, 263974, 264558
2768	94322238 (5535, 5536)	Novel Protein sim. GBank gi 5441322 emb CAB46721.1  - (AL031427) dJ167A19.1 (novel protein) [Homo sapiens]		UNCLASSIFIED	264488, 263994, 264489, 35696286, 22278998, 22278999, 264259, 29331822, 60432289, 29331826, 35696052, 264508, 264509, 264905, 264908, 264907, 29331830, 264908, 264909, 264510, 284511, 264512, 265008, 264910, 285009, 284591, 264592, 264593, 33657402, 264594, 284595, 264757, 264596, 265011, 264604, 265019, 264760, 264681, 18108351, 264764, 264288, 264766, 264688, 264768, 18108357, 264769, 264689, 21906766, 21906768, 21906769, 35695917, 265021, 265022, 33657023, 33657109, 264628, 264629, 18108374, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264555, 264638, 264639, 83373044, 264563, 264564, 264565, 264566, 264486, 264567

2769	95311088 (5537, 5538)	Novel Protein sim. GBank gi 5419859 emb CAB46375.1  - (AL096725) hypothetical protein [Homo sapiens]	tubulin	264488, 56182575, 264259, 29331822, 29331824, 66714117, 29331825, 60432289, 29331828, 264908, 29331830, 56182435, 264592, 33657402, 264448, 264369, 264288, 60170615, 264691, 33657023, 264692, 33657109, 18108374, 55811576, 284634, 264636, 56182323, 83373044, 60432113, 18108398, 22278995, 22278996, 22278999, 284105, 265006, 265019, 18108351, 264687, 21906765, 18108364, 264629, 18108374, 264631, 18108385, 18108388
2770	87730182 (5539, 5540)	Novel Protein sim. GBank gi 5701965 emb CAB52157.1  - (AL109736) WD repeat protein [Schizosaccharomyces pombe]	kinase	Contains protein domain (PF00400) - WD domain, G-beta repeat
2771	88084071 (5541, 5542)	Novel Protein sim. GBank gi 3093433 (AC004125) - Unknown gene product [Homo sapiens]	UNCLASSIFIED	264259, 29331822, 29331824, 29331825, 264359
2772	95357309 (5543, 5544)	Novel Protein sim. GBank gi 488553 ref NP_005465.1 pNY C - histone deacetylase 5	histone	264259, 29331822, 29331824, 66714117, 60432289, 29331827, 264905, 265009, 264592, 55812038, 65274444, 264766, 21806769, 33657109, 263978, 264555, 264638, 264557, 83373044, 264563, 264564, 264486
2773	94138984 (5545, 5546)	Novel Protein sim. GBank gi 3288888 (AC005253) - R26445_1 [Homo sapiens]	UNCLASSIFIED	18108374, 264686, 264687, 263976, 56182435, 264689, 55810784, 21905766, 35696423, 55811576, 65274791, 56181686, 55811957, 35695855, 264110, 265021, 264112, 265022, 265006, 265008, 264092, 264094, 60431850, 264637, 264691, 264259, 264692, 263981, 264594, 60433356, 56182323, 264693, 264757, 56182181, 55812038, 29331825, 60424269, 18108385, 29331826, 29331827, 27486261, 20331828, 35696052, 55811386, 264107, 60432113, 265017, 55811150, 18108351, 264681, 264906, 18108370, 264484, 264682, 20281069, 264448, 66712502, 264683, 264764, 264288, 264684, 264766, 263974
2774	87819808 (5547, 5548)	Novel Protein sim. GBank gi 465852 sp P34388 YLS3_CAEEL - HYPOTHETICAL 70.7 KD PROTEIN F09G8.3 IN CHROMOSOME III	ribosomalprot	22278995, 35696288, 22278996, 22278999, 264259, 29331826, 60432289, 35696052, 264112, 33657402, 21906754, 87168559, 265017, 265018, 264288, 21906766, 21906767, 21906769, 35695917, 265020, 265021, 33657023, 33657109, 18108370, 263976, 35696423, 35695855, 87168518, 22279000, 264482

2775	95307987 (5549, 5550)	Novel Protein sim. GBank gi 4688132 gb AAD27775.1 AF07704 - (AF077042) 30S ribosomal protein S7 homolog [Homo sapiens]	Contains protein domain (PF00177) - Ribosomal protein S7p/S5e	ribosomalprot	264488, 22278995, 56994075, 22278998, 3569288, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331827, 29331828, 35696052, 265007, 21908754, 265017, 265019, 264448, 264682, 264369, 264288, 18108354, 52644229, 264689, 21908765, 21908766, 21908767, 21908768, 21908769, 35695917, 265021, 265022, 60170815, 264691, 18108370, 35696423, 65274791, 35695855, 264634, 60431850, 60170394, 56182323, 264556, 18108388, 22279000, 264563, 264565, 264486, 264567
2776	87791557 (5551, 5552)				56182575, 22278998, 22278999, 264259, 29331822, 29331824, 264908, 29331830, 264510, 33657402, 21908754, 55811386, 265017, 265018, 264448, 264288, 21908765, 21908768, 21908767, 21908768, 21908769, 265020, 265021, 265022, 60170615, 55810764, 55811576, 264555, 56526486, 22279000
2777	79818729 (5553, 5554)			UNCLASSIFIED	264907, 264786
2778	82112411 (5555, 5556)			UNCLASSIFIED	264907, 264593, 264760, 264628
2779	87649729 (5557, 5558)	Novel Protein sim. GBank gi 4680711 gb AAD27745.1 AF13297 - (AF132970) CGI-36 protein [Homo sapiens]		UNCLASSIFIED	22278997, 264259, 29331824, 66714117, 35696052, 265006, 264512, 264448, 264288, 29148627, 18108364, 20281149, 18108370, 264629
2780	94679397 (5559, 5560)	Novel Protein sim. GBank gi 4758524 ref NP_004825.1 pHCK - HPK/GCK-like kinase	Contains protein domain (PF00780) - CIN domain	kinase	29146499, 65274791, 264634, 264639
2781	91220057 (5561, 5562)	Novel Protein sim. GBank gi 469352 gb AAD21222  - (AF069502) ubiquitin specific protease UBP43 [Mus musculus]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	60424179, 29331824, 60424269, 66714117, 29331826, 56182435, 87168474, 285017, 264764, 56181562, 21908765, 21908768, 21908769, 35695917, 265020, 33657023, 35695855, 56182323, 87168518
2782	94233146 (5563, 5564)	Novel Protein sim. GBank gi 4505013 ref NP_002310.1 pLRN - leucine-rich neuronal protein	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	65274572, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331826, 60432289, 29331828, 264905, 264907, 264908, 264909, 52644045, 265009, 265017, 265018, 264804, 265019, 264760, 264683, 264288, 264786, 264685, 264688, 264768, 52644229, 264689, 21908768, 265020, 265021, 264691, 18108362, 264692, 33657023, 264693, 33657109, 33657349, 18108370, 264628, 263978, 35695855, 264557, 56182323, 83373044, 18108385

2783	80016829 (5565, 5568)	Novel Protein sim. GBank gil728831 sp P39188 ALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY III		tm7	264909, 264628, 263978, 263981
2784	87614360 (5567, 5568)				264259, 29331822, 29331824, 29331825, 264482
2785	88071930 (5569, 5570)	Novel Protein sim. GBank gil2134933 pir [S58890 - collapsin response mediator protein - human		UNCLASSIFIED	22278999, 264908, 264758, 265018, 264769, 21906765, 21906768, 21906769, 265020, 264564
2786	87408542 (5571, 5572)	Novel Protein sim. GBank gil2073564 (U80223) - eukaryotic initiation factor eIF-2 alpha kinase; DGCN2 [Drosophila melanogaster]	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	kinase	284905, 265017
2787	87801268 (5573, 5574)	Novel Protein sim. GBank gil5174507 ref NP_006020.1 pMA1 - paraneoplastic neuronal antigen		UNCLASSIFIED	264768, 21906765, 21906768, 55811957, 22278999, 264093, 60170615, 264259, 29331822, 18108365, 29331824, 33857109, 29331827, 35696052, 264100, 264105, 264908, 263977, 55811576, 264635, 264637, 60433438, 60432113, 265017, 265019, 22279002, 55811150, 264369, 264288
2788	88080844 (5575, 5576)	Novel Protein sim. GBank gil3252826 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	
2789	85491275 (5577, 5578)	Novel Protein sim. GBank gil2485729 sp Q92556 Y281_HUMAN - HYPOTHETICAL PROTEIN KIAA0281 (HA8725)		UNCLASSIFIED	264103, 21906769, 264693
2790	87602784 (5579, 5580)	Novel Protein sim. GBank gil5101772 emb CAB45135.1  - (AJ242978) p821 [Homo sapiens]		struct	264488, 264769, 264910, 264631, 264636, 264690, 264691, 264259, 264638, 29331824, 264508, 264509, 264905, 264563, 264908, 264628, 18108370, 264907, 264764, 264908, 264288, 264909
2791	88083195 (5581, 5582)	Novel Protein sim. GBank gil2911266 (AC002550) - Unknown gene product [Homo sapiens]			21906764, 18108368
2792	95083783 (5583, 5584)	Novel Protein sim. GBank gil2854163 gb AAC02581.1  - (AF045642) No definition line found [Caenorhabditis elegans]			22278996, 22278997, 264259, 29331822, 29331824, 29331825, 66714117, 29331826, 60432289, 29331827, 35696052, 264908, 66712502, 29331830, 264909, 60432229, 60433356, 60433438, 33109954, 265010, 265011, 265017, 265018, 265019, 264682, 264448, 264288, 264369, 264766, 52644229, 21906765, 21906766, 265020, 265021, 33657023, 263974, 18108374, 65274781, 35695855, 264636, 264558, 264558, 56182323, 83373044, 18108385, 56526486, 60432113, 22279000, 264567
2793	87425478 (5585, 5586)			UNCLASSIFIED	264259, 60432289, 66712502, 265009, 264636
2794	85794830 (5587, 5588)			UNCLASSIFIED	264689, 265022, 264691, 18108388, 264567

2785	95334888 (5589, 5590)	Novel Protein sim. GBank gi 5454146 ref NP_006348.1 pUBE2 - ubiquitin-conjugating enzyme E2E 3 (homologous to yeast UBC4/5)	Contains protein domain (PF00179) - Ubiquitin-conjugating enzyme	ubiquitin	65274572, 56182575, 35696286, 22278998, 22278999, 60432049, 264259, 29331822, 29331825, 66714117, 60432289, 29331828, 29331827, 35696052, 29331828, 264907, 66712502, 56182435, 264511, 265007, 264512, 264910, 60432229, 60433356, 60433438, 33109954, 85658542, 265018, 265019, 264288, 264686, 21906764, 21906765, 21906768, 21906769, 21906768, 55811957, 265020, 265021, 285022, 52644150, 33657023, 264693, 65274620, 33657109, 35696423, 55811576, 65274791, 56182323, 56526486, 60432113, 22279002, 264482, 264563, 264484, 264567
2786	94848857 (5591, 5592)	Novel Protein sim. GBank gi 4680651 gb AAD27715.1 AF13294 - (AF13294) CGI-06 protein [Homo sapiens]		UNCLASSIFIED	18108394, 65274572, 56182575, 56994075, 22278999, 264480, 60432049, 264259, 29331822, 29331824, 29331826, 35696052, 264509, 264908, 264907, 264908, 66712502, 56182435, 264510, 265006, 264512, 265007, 265008, 264910, 265008, 264591, 264592, 60433356, 264594, 264595, 55812038, 264596, 21906754, 60174639, 87168474, 265010, 265011, 265017, 265018, 265019, 55811150, 264762, 18108351, 264448, 264682, 264369, 264288, 264685, 264766, 264687, 56181562, 264769, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 52644150, 264692, 33657023, 18108362, 264693, 65274620, 33657109, 33657182, 27486285, 33657349, 18108374, 35696423, 65274781, 35695855, 264556, 264557, 56182323, 264558, 60170394, 83373044, 65274727, 87168518, 22279000, 264563, 264564, 264565, 264566, 264567
2787	95110780 (5593, 5594)	Novel Protein sim. GBank gi 4838557 gb AAD31040.1  - (AF143859) potassium channel modulatory factor DEBT-91 [Mus musculus]	Contains protein domain (PF00569) - Zinc finger present in dystrophin, CBP/p300	UNCLASSIFIED	55182575, 22278995, 22278996, 22278997, 29331827, 29146499, 264509, 264906, 56182435, 264757, 21906754, 265010, 265017, 265018, 264681, 264682, 264683, 264686, 21906765, 21906767, 21906768, 21906769, 29148629, 265020, 52644150, 264690, 33657182, 264629, 18108376, 56182323, 22278002, 264563
2788	86198005 (5595, 5596)	Novel Protein sim. GBank gi 2852645 (AF007160) - unknown [Homo sapiens]		UNCLASSIFIED	264488, 264490, 264259, 264448, 20281149, 20281152, 264556, 264557, 264558, 264559, 264483, 264486, 264567

2799	88090651 (5597, 5598)	Novel Protein sim. GBank gij252825 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	22278996, 22278997, 22278998, 22278999, 60432048, 264259, 29331826, 35698052, 29331828, 264809, 60433356, 33657402, 33109954, 87168474, 264448, 52644229, 21906766, 21906767, 21906768, 35695917, 265020, 265021, 265022, 52644150, 35695855, 264634, 60432113, 22279000
2800	88318481 (5599, 5600)	Novel Protein sim. GBank gij240301[dijBAA74929.1] - (AB020713) KIAA0906 protein [Homo sapiens]	glycoprotein		264488, 56894075, 264259, 20281099, 29331825, 29331827, 264905, 56182435, 255006, 265011, 87168559, 265017, 265019, 264448, 264288, 264766, 264686, 60170815, 264691, 264692, 27486265, 264628, 264629, 264636, 264557, 264558, 264559, 87168518, 264584, 264586, 264567
2801	86068814 (5601, 5602)		Contains protein domain (PF00627) - UBA domain		265007, 264687
2802	88082477 (5603, 5604)	Novel Protein sim. GBank gij2337865 (AC002464) - organic cation transporter; 50% similarity to JC4884 (PID:g2143892) [Homo sapiens]	Contains protein domain (PF00083) - Sugar (and other) transporter	transport	264448, 35695855
2803	79577446 (5605, 5606)			UNCLASSIFIED	264639
2804	57111131 (5607, 5608)	Novel Protein sim. GBank gij4559368[gb AAD23028.1 AC00658 - (AC006585) hypothetical protein [Arabidopsis thaliana]	Contains protein domain (PF01585) - G-patch domain	peptidase	264566
2805	87398486 (5609, 5610)				264092, 264259, 29331822, 29331824, 29331827, 29331828, 264508, 265007, 265009, 21906754, 264369, 264288, 264632, 60170394, 264563, 264482
2806	87888951 (5611, 5612)	Novel Protein sim. GBank gij1168973[sp P44403 CLPB_HAEIN - CLPB PROTEIN		UNCLASSIFIED	22278995, 22278996, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29146498, 265008, 265009, 60433438, 265017, 265018, 265019, 264448, 264288, 21906765, 21906767, 21906769, 29148629, 265022, 52644150, 56182323



2807	91720702 (5613, 5614)	Novel Protein sim. GBank gi 4468310 emb CAB37991  - (AL031432) dJ465N24.1 (PUTATIVE novel protein similar to predicted yeast and worm proteins) [Homo sapiens]		UNCLASSIFIED	52644507, 52645156, 52646842, 18108398, 56182575, 22278994, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 35696052, 29331828, 33656970, 264100, 264105, 264907, 52644045, 60433356, 264594, 60433438, 52646317, 21908754, 33109954, 33657084, 52644296, 87188474, 265011, 87188559, 265017, 265018, 265019, 18108331, 264448, 264763, 264687, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 265022, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486264, 27486285, 35695763, 18108370, 18108372, 18108374, 18108376, 263977, 18108377, 35696423, 35695855, 52644332, 83373044, 18108385, 18108387, 87168518, 60432113
2808	95359111 (5615, 5616)	Novel Protein sim. GBank gi 5541863 emb CAB51071.1  - (AL096857) hypothetical protein [Homo sapiens]		MHC	60432289, 264510, 265010, 265017, 265018, 264681, 264686, 265021, 264690, 22279000, 264566
2809	86083530 (5617, 5618)	Novel Protein sim. GBank gi 2772561 (AC004002) - similar to ciliary dynein beta heavy chain; 78% Similarity to P23098 (PID:gi118965) [Homo sapiens]		ATPase_associated	18108331
2810	87259032 (5619, 5620)				
2811	91235845 (5621, 5622)			UNCLASSIFIED	264569, 22278996, 264091, 264259, 29331828, 29146499, 29146629, 29148784, 264693
2812	86093334 (5623, 5624)	Novel Protein sim. GBank gi 3264583 (AC005189) - match to ESTs H97758 (NID:gi118843) and AA085546 (NID:gi1828773) [Homo sapiens]		UNCLASSIFIED	22278999, 264259, 66712502, 264693
2813	91218755 (5625, 5626)	Novel Protein sim. GBank gi 4240273 dbj BAA74915.1  - (AB020699) KIAA0892 protein [Homo sapiens]		UNCLASSIFIED	264108
2814	90980906 (5627, 5628)	Novel Protein sim. GBank gi 3548791 (AC005620) - R33590_1 [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	transcriptfactor	56182575, 29331828, 264906, 66712502, 55811386, 265017, 265018, 265019, 264683, 265020, 87168518, 60432113
2815	79774521 (5629, 5630)			UNCLASSIFIED	65274572
2816	95358228 (5631, 5632)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1  - (AJ243460) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264907, 264909
				UNCLASSIFIED	264488, 35696286, 29331825, 29331828, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 265011, 264760, 264681, 264766, 264769, 264689, 21906765, 264693, 264628, 18108370, 264629, 264631, 264634, 264563, 264564, 264566, 264486

2817	87749542 (5633, 5634)	Novel Protein sim. GBank gi 1293846 (U56966) - coded for by C. elegans cDNA yk30b3.5; coded for by C. elegans cDNA yk30b3.3 [Caenorhabditis elegans]			264259, 29331822, 29331827, 264508, 284309, 264905, 264907, 264908, 264909, 56182435, 264510, 265006, 264511, 264512, 264593, 264758, 265010, 264760, 264761, 264764, 264288, 264687, 264769, 55811957, 35695917, 33657109, 263978, 264634, 264636, 264639, 264584, 264565, 264566, 264486, 264587
2818	86073579 (5635, 5636)	Novel Protein sim. GBank gi 549986 (U13149) - possible asporary-associated protein [Pennisetum ciliare]	Contains protein domain (PF00023) - Antk repeat	UNCLASSIFIED	66712502
2819	87793527 (5637, 5638)			UNCLASSIFIED	264909, 264511
2820	87765744 (5639, 5640)	Novel Protein sim. GBank gi 4929773 gb AAD34147.1 AF15209 - (AF152097) CGI-05 protein [Homo sapiens]	Contains protein domain (PF00919) - Uncharacterized protein family UPF0004	UNCLASSIFIED	18108394, 52646365, 52644045, 264112, 265009, 21906754, 265017, 18108351, 264683, 264389, 264686, 264769, 21906769, 33657023, 33657109, 18108370, 18108374, 18108377, 18108385
2821	95320511 (5641, 5642)	Novel Protein sim. GBank gi 399144 sp P02747 C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264636
2822	94260221 (5643, 5644)	Novel Protein sim. GBank gi 2224671 dbj BAA20820  - (AB002363) KIAA0365 [Homo sapiens]		UNCLASSIFIED	264259, 29331822, 60432289, 264905, 60433356, 33657402, 265011, 264760, 21906765, 60170815, 264638
2823	95320513 (5645, 5646)	Novel Protein sim. GBank gi 399144 sp P02747 C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264766
2824	95320515 (5647, 5648)	Novel Protein sim. GBank gi 399144 sp P02747 C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR		UNCLASSIFIED	264907
2825	19742170 (5649, 5650)				264760
2826	94311905 (5651, 5652)	Novel Protein sim. GBank gi 3858683 emb CAA22020  - (AL0333503) conserved hypothetical protein [Candida albicans]			52645156, 22278994, 35696286, 22278997, 22278998, 52645080, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 33656970, 52644045, 52646317, 33657084, 52644296, 265017, 265018, 265019, 264288, 21906764, 21906765, 21906766, 21906767, 21906769, 35695917, 52644150, 33657109, 33657182, 27486261, 27486262, 33657349, 27486265, 35695763, 35696423, 35695855, 52644332, 18108385, 87168518, 264484

2827	95320519 (5653, 5654)	Novel Protein sim. GBank gi 399144 sp P02747 C1QC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264488, 263994, 264489, 65274572, 29331822, 66714117, 29331827, 29331828, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 285006, 264511, 265007, 264512, 265008, 264910, 264591, 60432229, 264592, 264593, 264595, 264596, 21906754, 265011, 264600, 264601, 264602, 265017, 264604, 264605, 264288, 264766, 264767, 264689, 55811957, 264534, 264690, 264691, 264692, 264693, 263972, 264629, 35695855, 264631, 264634, 264635, 264555, 264636, 264637, 264556, 264638, 264558, 83373044, 18108385, 60432113, 22279002, 264563, 264564, 264565, 264566, 264567, 264486, 18108391
2828	91228615 (5655, 5656)	Novel Protein sim. GBank gi 3598874 (AF077000) - protein tyrosine phosphatase TD14 [Rattus norvegicus]	Contains protein domain (PF00102) - Protein-tyrosine phosphatase	phosphatase	29331822, 35696052, 264104, 264908, 265007, 264591, 265010, 265011, 265019, 264766, 264686, 55811957, 18108370, 18108374, 55810764, 35696123, 55811576, 56182323, 83373044, 87168518 22278996, 22278997, 264091, 264093, 60432049, 264259, 29331822, 29331825, 29331827, 29331828, 264905, 264509, 66712502, 264510, 264511, 264593, 60433438, 21906754, 265011, 264603, 18108351, 264288, 21906765, 21906768, 21906769, 29148629, 52644150, 264693, 33657109, 18108374, 264634, 18108385, 60432113, 22279000, 264565, 264486 265008, 265019, 264639, 22279002
2829	87651244 (5657, 5658)	Novel Protein sim. GBank gi 4680688 gb AAD27734.1 AF13295 - (AF132959) CGI-25 protein [Homo sapiens]			
2830	88087109 (5659, 5660)	Novel Protein sim. GBank gi 2498667 sp Q61200 NPH1_MOUSE - NEUREXOPHILIN 1			
2831	87614717 (5661, 5662)			UNCLASSIFIED	265017
2832	87631809 (5663, 5664)			UNCLASSIFIED	22278997, 22278999, 52646317, 264288, 264688, 21906767, 60431528, 264638, 22279000
2833	87612938 (5665, 5666)	Novel Protein sim. GBank gi 5262615 emb CAB45747.1  - (AL080158) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264555, 264556, 264558
2834	86974703 (5667, 5668)	Novel Protein sim. GBank gi 2224567 dbj BAA20772  - (AB002311) KIAA0313 [Homo sapiens]			263972
2835	87775712 (5669, 5670)	Novel Protein sim. GBank gi 4589532 dbj BAA76788.1  - (AB023161) KIAA0944 protein [Homo sapiens]		ATPase-associated	60432289, 29331828, 265008, 265010, 265017, 264448, 55811957, 265020, 18108370
2836	85724748 (5671, 5672)	Novel Protein sim. GBank gi 2351568 (U76618) - N-RAP [Mus musculus]	Contains protein domain (PF00412) - LIM domain containing proteins	transcript factor	264259, 264112, 265010, 264762, 264764, 293974, 264555, 264558, 264559

2837	87766482 (5673, 5674)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	18108394, 22278897, 22278998, 29331822, 264907, 264909, 265006, 265007, 265018, 265019, 264682, 264288, 21908766, 21906767, 55811957, 35695917, 18108374, 56182323, 22279000, 22279002
2838	87775392 (5675, 5676)	Novel Protein sim. GBank gi 973378 (U31263) - core protein [Hepatitis C virus]		UNCLASSIFIED	18108394, 18108397, 264259, 29331826, 265007, 265019, 264448, 18108368, 20281149, 264565, 264567
2839	85799317 (5677, 5678)	Novel Protein sim. GBank gi 1575515 (U64899) - thrombospondin-related anonymous protein [Plasmodium gallinaceum]		UNCLASSIFIED	264555
2840	87774665 (5678, 5680)	Novel Protein sim. GBank gi 2224605 dbj BAA20790  - (AB002330) KIAA0332 [Homo sapiens]			264509, 264511, 265011, 264288, 264769, 265020, 264634, 264636, 264556
2841	86982568 (5681, 5682)	Novel Protein sim. GBank gi 2224605 dbj BAA20790  - (AB002330) KIAA0332 [Homo sapiens]	Contains protein domain (PF000076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	56182575, 35696052, 264907, 264908, 264909, 264910, 264593, 264595, 264766, 265022, 264691, 33657182, 35695763, 18108370, 35695855, 264631, 264559, 264563, 264567
2842	80080086 (5683, 5684)	Novel Protein sim. GBank gi 5578957 emb CAB51350.1  - (AL050306) dJ47587.1 (novel KIAA0615 and KIAA0323 LIKE protein) [Homo sapiens]		UNCLASSIFIED	264600
2843	91012494 (5685, 5686)	Novel Protein sim. GBank gi 5578957 emb CAB51350.1  - (AL050306) dJ47587.1 (novel KIAA0615 and KIAA0323 LIKE protein) [Homo sapiens]		UNCLASSIFIED	264906, 264907, 264908, 264909, 264910, 264764, 35695855, 83373044, 18108385
2844	56731154 (5687, 5688)	Novel Protein sim. GBank gi 585123 sp Q08878 FBLC_MOUSE - FIBULIN-1, ISOFORM C PRECURSOR (BASEMENT-MEMBRANE PROTEIN 90) [BM-90]	Contains protein domain (PF00008) - EGF-like domain		264685
2845	84321719 (5689, 5690)	Novel Protein sim. GBank gi 5420387 emb CAB46679.1  - (AJ243459) proteophosphoglycan [Leishmania major]	homeobox		29146498, 87168474, 264686, 35696423, 83373044, 264564
2846	88318613 (5691, 5692)	Novel Protein sim. GBank gi 5306263 gb AAD41995.1 AC006233) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	29331830, 264909, 265008, 265011, 87188559, 264629, 264556
2847	81811757 (5693, 5694)	Novel Protein sim. GBank gi 3399676 (AC005390) - R31180.1 [Homo sapiens]		UNCLASSIFIED	264908, 264766, 264769, 264629, 264637, 264566
2848	87612943 (5695, 5696)	Novel Protein sim. GBank gi 5262615 emb CAB45747.1  - (AL080156) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264490, 264259, 264508, 264905, 264907, 264510, 265007, 265008, 264591, 264592, 264593, 264594, 264595, 55812038, 265010, 265011, 264604, 264763, 264764, 264765, 264766, 264686, 264628, 264829, 264555, 264636, 264556, 264557, 264638, 264558, 264559, 264563, 264566, 264567
2849	88084283 (5697, 5698)	Novel Protein sim. GBank gi 3342218 (AC004131) - Unknown gene product [Homo sapiens]			29331822, 35696052, 264509, 264906, 265007, 264594, 265018, 264288, 263972, 35696423, 18108384, 56526486, 18108390
2850	87623636 (5699, 5700)			UNCLASSIFIED	35696286, 265008, 265009, 265018, 264288, 35695917, 264693, 18108374, 35695855, 22279000

2851	87820548 (5701, 5702)	Novel Protein sim. GBank gjk4321619[gb AAD15768.1] - (AF051098) seven transmembrane domain orphan receptor [Mus musculus]		UNCLASSIFIED	264906, 264907, 56182435, 264758, 55811386, 265010, 18108351, 264448, 264389, 21906765, 264691, 264692, 264693, 35695855, 264556, 18108385, 264567, 264591
2852	86987023 (5703, 5704)	Novel Protein sim. GBank gjl1825729 (U88308) - similar to drosophila membrane protein PATCHED SP:P18502 (PID:g129645) [Caenorhabditis elegans]		UNCLASSIFIED	
2853	87784630 (5705, 5706)	Novel Protein sim. GBank gjl2702347 (AF027503) - putative membrane-associated guanylate kinase 1 [Mus musculus]	Contains protein domain (PF00397) - WW domain	kinase	56182575, 55811150, 264690, 27486262, 27486265, 264632, 56182323, 56526486, 22279002
2854	88083557 (5707, 5708)	Novel Protein sim. GBank gjl2785825 (AC004021) - kelch protein, ring canal component involved in cytoplasmic bridges; 77% Similarity to A45773 (PID:g1079096) [Homo sapiens]	Contains protein domain (PF01344) - Kelch motif	dna_ma_bind	35696286, 29331824, 28331826, 29331828, 264908, 264766, 264693, 22279002, 264482
2855	94723856 (5709, 5710)	Novel Protein sim. GBank gjl1504040[gb BAA13218] - (D86983) similar to D.melanogaster peroxidase(U11052) [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	22278994, 29331822, 29331824, 29331825, 264906, 264908, 265009, 33109954, 265018, 265019, 264448, 21906765, 265020, 264690, 27486265, 83373044, 22279000, 22279002, 264482
2856	88093359 (5711, 5712)	Novel Protein sim. GBank gjl3264583 (AC005189) - match to ESTs H97758 (NID:g1118643) and AA085548 (NID:g1628773) [Homo sapiens]			21906766, 22278997, 265022, 29331822, 29331826, 27486262, 265007, 265009, 265017, 264482, 264563, 18108351
2857	95348286 (5713, 5714)	Novel Protein sim. GBank gjl3041855 (AC004537) - similar to tumor suppressor p33ING1; similar to AF044076 (PID:g2829208) [Homo sapiens]	Contains protein domain (PF00628) - PHD-finger	struc	22278995, 35696286, 29331824, 29331825, 35696032, 264103, 264108, 56182435, 21906765, 21906769, 265020, 18108368, 35695763, 22279002, 264563
2858	87434748 (5715, 5716)	Novel Protein sim. GBank gjl462585[sp P35227]ME18_HUMAN - DNA-BINDING PROTEIN MEL-18 (ZINC FINGER PROTEIN 144)	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	dna_ma_bind	264569, 264887, 22278995, 22278996, 22278997, 22278999, 264259, 29331826, 29331827, 29331828, 264509, 264905, 264906, 29331830, 264908, 52644045, 264909, 264511, 264512, 265007, 265008, 264910, 265009, 264593, 60433356, 264595, 264758, 21906754, 265010, 265011, 264604, 265018, 264760, 18108351, 264763, 264682, 264764, 264765, 264288, 264369, 264685, 264766, 264788, 18108357, 264769, 21906766, 21906767, 265021, 264534, 60170615, 264691, 264692, 18108370, 264629, 18108374, 264631, 264636, 263981, 18108381, 264558, 18108385, 22279002, 264564, 264566, 264486, 264567
2859	90937675 (5717, 5718)	Novel Protein sim. GBank gjk4325320[gb AAD17331.1] - (AF124427) claudin-15 [Mus musculus]		UNCLASSIFIED	60424179, 65274572, 29331828, 264905, 264511, 264758, 265011, 21906767, 21906769, 55811957, 265021, 56182323

2860	87532589 (5719, 5720)	Novel Protein sim. GBank gij4469186jemb[CAB38414.1] - (AL031588) dJ1163J1.2.1 (novel protein similar to C. elegans B0035.16 and bacterial RNA (5-Methylaminomethyl-2-thiouridylyl)-Methyltransferases) (isoform 1) [Homo sapiens]		UNCLASSIFIED	284259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 284510, 284511, 33109954, 18108351, 264683, 264765, 264369, 264686, 21906765, 264691, 264692, 284693, 18108388, 22279002, 264482
2861	86698507 (5721, 5722)	Novel Protein sim. GBank gij3941730 (AF108083) - BS4 [Homo sapiens]			284368, 264692
2862	87569585 (5723, 5724)	Novel Protein sim. GBank gij4505013jref[NP_002310.1]pLRN] - leucine-rich neuronal protein		UNCLASSIFIED	264691, 264638
2863	91220421 (5725, 5726)	Novel Protein sim. GBank gij3249088 (AC004473) - Contains similarity to goliath protein gblM97204 from D. melanogaster. [Arabidopsis thaliana]	Contains protein domain (PF00096) - struct Zinc finger, C2H2 type		56994075, 35696286, 22278998, 29331822, 29331824, 35696052, 29331828, 284106, 284511, 55812038, 33657084, 55811386, 265018, 265019, 21906765, 21906766, 21906769, 35695917, 265020, 265022, 33657023, 33657109, 33657349, 264629, 18108376, 60431850, 58182323, 18108385, 18108387, 87168518, 22279002
2864	87420030 (5727, 5728)	Novel Protein sim. GBank gij1079451pirj[A55463] - tropomodulin, skeletal muscle - chicken	struct		284259, 284910
2865	95312181 (5729, 5730)	Novel Protein sim. GBank gij438840 (L19048) - MSA-2 [Plasmodium falciparum]			22278995, 21906764, 264482
2866	95105480 (5731, 5732)	Novel Protein sim. GBank gij585703spjQ07066[PMP2_RAT] - 22 KD PEROXISOMAL MEMBRANE PROTEIN	glycoprotein		65274572, 22278996, 22278998, 22278999, 284259, 29331824, 29331827, 29331828, 60433438, 21906754, 265018, 264448, 284764, 52844150, 83373044
2867	86808001 (5733, 5734)	Novel Protein sim. GBank gij4580897gbjAAD24571.1[AF12108] cAMP inducible 2 protein [Mus musculus]	UNCLASSIFIED		264488, 284768, 21906768, 22278998, 265022, 284259, 284508, 284905, 264907, 284511, 284910, 264635, 264636, 264637, 265011, 265017, 265018, 285018, 264583, 264088, 264566, 284764, 264369, 264567, 264488, 264288, 264768

2868	95303283 (5735, 5736)	Novel Protein sim. GBank g 1292868 emb CAA63923  - (X94232) t-Cell activation protein [Homo sapiens]		18108392, 264488, 22278994, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 264508, 52644045, 264828, 265006, 265007, 265008, 285009, 264591, 60432229, 264593, 60433356, 264595, 21906754, 265017, 265019, 264682, 264369, 21906765, 21906766, 21906767, 21906768, 265021, 265022, 264691, 33657182, 18108388, 27486261, 27486262, 27486264, 27486265, 18108370, 18108374, 35696423, 35695855, 264632, 5618233, 87188518, 264404, 22279000, 22279002, 264482, 264563, 264564, 264567, 264487
2869	88094412 (5737, 5738)		UNCLASSIFIED	264369
2870	84404574 (5739, 5740)		UNCLASSIFIED	264305, 264908, 264764, 21906769, 264634
2871	88318621 (5741, 5742)	Novel Protein sim. GBank g 5306263 gb AAD41995.1 AC006233 - (AC006233) unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264259, 29331822, 60432289, 29331827, 264907, 265008, 265017, 265018, 264682, 264764, 18108354, 265021, 27486285, 264629, 18108387, 264567
2872	95312187 (5743, 5744)	Novel Protein sim. GBank g 112205 pir B39066 - proline-rich protein 15 - rat	kinase	263981
2873	88094252 (5745, 5746)		UNCLASSIFIED	264488, 18108374, 264768, 264687, 264688, 264689, 35696423, 35696286, 35695917, 264510, 264511, 265007, 264512, 265008, 264910, 264534, 264634, 264635, 264555, 264592, 264259, 264558, 60433438, 60432289, 35696052, 265011, 264600, 264601, 60432113, 264508, 264563, 264482, 264509, 264805, 264906, 264564, 18108351, 264763, 18108370, 264907, 264566, 264908, 264764, 264288, 264567, 264909, 264486, 264768, 18108391

2874	94313548 (5747, 5748)	Novel Protein sim. GBank gij212854 (AC004005) - unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	52644507, 52645156, 52646365, 56182575, 22278994, 22278995, 56994075, 35696286, 60432049, 264259, 52645080, 29331822, 56182181, 29331824, 60424269, 29331825, 66714117, 29331826, 29331827, 35696052, 29331828, 33656970, 66712502, 52644045, 265007, 265009, 60433356, 264758, 55812038, 18108348, 52646317, 33109954, 33657084, 265017, 264604, 265018, 265019, 264682, 264369, 264288, 264688, 52644229, 21906766, 21906767, 21906768, 55811957, 265020, 265021, 265022, 52644150, 33657023, 52645129, 18108374, 18108376, 35696423, 56182323, 18108387, 87168518, 60432113, 22279000, 22279002, 264563, 264565
2875	88083726 (5749, 5750)	Novel Protein sim. GBank gij2781386 (AC004010) - similar to Leucine-rich transmembrane proteins; 44% similarity to U42767 (PID:g1736918) [Homo sapiens]	Contains protein domain (PF00560) - glycoprotein Leucine Rich Repeat	22278996, 22278997, 22278999, 29331826, 29331828, 29148499, 66712502, 265008, 265017, 18108351, 264683, 264689, 21906767, 18108376, 18108377, 55811576, 60170394, 22279000, 264487
2876	88090854 (5751, 5752)	Novel Protein sim. GBank gij2979530 (AC004449) - R33883_2 [Homo sapiens]	Contains protein domain (PF00167) - fgl Fibroblast growth factor	
2877	94747029 (5753, 5754)	Novel Protein sim. GBank gij4704208 (emb CAB41646.1  - (AL035419) dJ1100H13.1 (putative novel protein) [Homo sapiens]		52646365, 65274572, 56182575, 22278997, 22278998, 22278999, 60432049, 52645080, 60424269, 60432289, 29331827, 35696052, 29331828, 66712502, 52644045, 56182435, 60433356, 33657402, 33657084, 265019, 55811150, 264448, 264369, 21906766, 21906768, 21906769, 265020, 33657023, 33657109, 33657182, 27486262, 264629, 60431528, 55811576, 52644332, 56182323, 264558, 83373044, 18108385, 56526486, 60432113, 22279000, 22279002, 264563
2878	88095309 (5755, 5756)	Novel Protein sim. GBank gij387675 (emb CAB03067  - (Z81077) predicted using GeneFinder. Similarity to Yeast protein 8248 (TR:G587531) [Caenorhabditis elegans]	UNCLASSIFIED Contains protein domain (PF01207) - Uncharacterized protein family UPF0034	264488, 264259, 29331822, 29331826, 264905, 264509, 264907, 264909, 264510, 265006, 264511, 264512, 33657402, 264758, 21906754, 18108351, 264681, 264682, 264288, 264684, 264685, 264768, 264689, 21906769, 264690, 33657023, 264693, 18108364, 33657109, 264629, 18108374, 264630, 264632, 264556, 264637, 264639, 264558, 18108385, 18108387, 264563, 264564, 264565, 264566, 264486, 264567



2879	87869122 (5757, 5766)	Novel Protein sim. GBank gi 4895145 gb AAD32752.1  - (AF127374) unknown [Streptomyces lavendulae]	Contains protein domain (PF00315) - Uracil-DNA glycosylase	UNCLASSIFIED	18108359, 264259, 264905, 18108370, 264629, 264908, 264909, 18108374, 18108377, 265006, 264910, 264637, 60170394, 264559, 265017, 264564, 264565, 264587, 264684, 264369
2880	94851439 (5759, 5760)	Novel Protein sim. GBank gi 4880703 gb AAD27741.1  (AF13296 - (AF132966) CGI-32 protein [Homo sapiens])			264488, 52646365, 52646642, 22278994, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 29331828, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 52644045, 264510, 265006, 264511, 265007, 264512, 265009, 264910, 264594, 21906754, 52646317, 52644296, 87188559, 264600, 264604, 264605, 264760, 264764, 264288, 264766, 264768, 264687, 264769, 21906766, 21906769, 35695917, 265021, 264690, 264692, 33657023, 52645129, 33657109, 33657182, 27486262, 33657349, 264629, 18108374, 35695855, 264634, 264635, 264636, 264637, 264638, 264557, 52644332, 264558, 264559, 83373044, 264404, 22279000, 264563, 264483, 264567, 264486
2881	87650539 (5761, 5762)	Novel Protein sim. GBank gi 733571 (U23452) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	22278998, 29331822, 52644045, 21906765, 264639, 60432113
2882	87714367 (5763, 5764)	Novel Protein sim. GBank gi 118112 (U41559) - No definition line found [Caenorhabditis elegans]			264488, 22278996, 22278999, 29331822, 29331826, 264908, 60170831, 60433356, 55812038, 264681, 264682, 264686, 264687, 264688, 21906768, 21906769, 264693, 263967, 18108374, 55811576, 56182323, 22279002, 264566
2883	95362875 (5765, 5766)	Novel Protein sim. GBank gi 4888008 gb AAD31087.1  (AF10693 - (AF106934) vitamin D receptor-interacting protein [Homo sapiens])	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	264112, 264682
2884	87784843 (5767, 5768)	Novel Protein sim. GBank gi 2224697 dbj BAA20832  - (AB002376) KIAA0378 [Homo sapiens]		UNCLASSIFIED	265018, 264634
2885	83006308 (5769, 5770)	Novel Protein sim. GBank gi 1255889 (U53344) - T07H6.5 gene product [Caenorhabditis elegans]		UNCLASSIFIED	264686, 264693
2886	91237823 (5771, 5772)	Novel Protein sim. GBank gi 3882323 dbj BAA34521.1  - (AB018344) KIAA0801 protein [Homo sapiens]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complementrecept	60432049, 264259, 28331828, 264908, 264511, 264595, 60433438, 264596, 265017, 264605, 263969, 263972, 264555, 83373044, 87188518, 264566
2887	91227860 (5773, 5774)	Novel Protein sim. GBank gi 3882323 dbj BAA34521.1  - (AB018344) KIAA0801 protein [Homo sapiens]		UNCLASSIFIED	18108351, 264688, 264629, 264631, 264639, 83373044, 264482
2888	95105816 (5775, 5776)	Novel Protein sim. GBank gi 4508015 ref NP_003447.1 pZNF2 - zinc finger protein 205	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264488, 264259, 29331828, 264508, 264906, 264593, 264758, 264766, 264769, 18108374, 83373044, 264486

2889	87606562 (5777, 5778)				UNCLASSIFIED	56994075, 29331824, 265009, 264760, 18108354, 264288
2890	78703853 (5778, 5780)	Novel Protein sim. GBank gl 854085 emb CAA583371 - (X83413) U88 [Human herpesvirus 6]			UNCLASSIFIED	264591, 264766
2891	88094428 (5781, 5782)	Novel Protein sim. GBank gl 3877750 emb CAB01508  - (Z78064) predicted using Genefinder; similar to collagen; cDNA EST EMBL:D65885 comes from this gene; cDNA EST EMBL:D68451 comes from this gene; cDNA EST EMBL:D66028 comes from this gene; cDNA EST EMBL:D69658 comes from this gene...			UNCLASSIFIED	264591, 264595, 264369, 264685, 264693, 264628, 264563, 264566
2892	95419745 (5783, 5784)	Novel Protein sim. GBank gl 4928759 gb AAD34140.1 AF15190 - (AF151903) CGI-145 protein [Homo sapiens]			UNCLASSIFIED	65274572, 35696286, 29331828, 264110, 265009, 60433438, 265018, 265019, 18108351, 264288, 21906765, 21906766, 21906769, 55811957, 264690, 65274620, 263967, 35695855
2893	87798014 (5785, 5786)				UNCLASSIFIED	264259, 265010, 264682, 18108370, 264555, 264556
2894	87755985 (5787, 5788)	Novel Protein sim. GBank gl 5669015 gb AAD46135.1  - (AF080171) zinc finger protein ZNF232 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type		transcript factor	264259, 265008, 60433438, 52644296, 265011, 264369, 35695917, 18108381, 18108382, 18108388
2895	86938778 (5789, 5790)	Novel Protein sim. GBank gl 3924708 emb CAA84646  - (Z35597) Weak similarity with sea squirt nidogen precursor protein (blastp score 71); cDNA EST EMBL:T02069 comes from this gene; cDNA EST EMBL:D76135 comes from this gene; cDNA EST EMBL:D73147 comes from this gene; cDNA EST EMB...	Contains protein domain (PF01437) - Plexin repeat			29331824, 265007, 264762, 264636, 264563
2896	87752122 (5781, 5792)	Novel Protein sim. GBank gl 4885549 ref NP_005456.1 ppKBC - protein kinase B gamma	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase		18108392, 18108394, 18108398, 22278998, 264259, 28331822, 29331824, 29146499, 264906, 264908, 265007, 265009, 265018, 265019, 264369, 264685, 264689, 21906766, 265021, 264693, 33657182, 264639, 18108384, 18108388, 264567
2897	95413057 (5793, 5794)	Novel Protein sim. GBank gl 4502877 ref NP_001296.1 pCLDN - Clostridium perfringens enterotoxin receptor 1			UNCLASSIFIED	60424179, 56182575, 22278996, 35696286, 22278997, 22278999, 60432049, 264259, 29331824, 29331825, 60424269, 60432289, 29331826, 29331828, 35696052, 264908, 56182435, 265009, 264910, 60170831, 60431735, 60433356, 60433438, 65274444, 55811386, 265018, 18108351, 264448, 264288, 264687, 21906765, 21906766, 21906767, 21906769, 55811957, 35695917, 264534, 33657023, 33657109, 35695763, 264628, 264629, 60431528, 18108374, 55810764, 55811576, 35696423, 35695855, 264555, 56182323, 18108385, 264404, 22279000, 22279002, 264566
2898	87750340 (5795, 5796)				UNCLASSIFIED	22278995, 264604, 18108385, 264566

2899	80357670 (5797, 5798)				UNCLASSIFIED	264764, 21906764, 264692
2800	94233538 (5799, 5800)	Novel Protein sim. GBank gi 4581470 emb CAB40137.1  - (Y18483) SLC7A8 protein [Homo sapiens]			glycoprotein	85274572, 56182575, 35696286, 60432049, 264259, 29331824, 66714117, 29331826, 35696052, 29331828, 66712502, 56182435, 285006, 265007, 265008, 265009, 60433356, 264758, 265018, 264764, 264765, 264288, 264768, 21906764, 21906768, 21906769, 265020, 264692, 264693, 32833986, 264631, 83373044, 60432113
2801	87444731 (5801, 5802)	Novel Protein sim. GBank gi 4759272 ref NP_004614.1 pTTC4 - tetralricopeptide repeat domain 4			phosphatase	22278995, 22278997, 22278999, 60432049, 29331822, 29331824, 29331825, 29331827, 35696052, 33656970, 264910, 265009, 21906754, 33657084, 87168474, 265010, 265018, 21906764, 21906765, 21906766, 21906767, 21906769, 33657023, 264693, 33657109, 33657349, 35696423, 35695855, 263981, 56182323, 22279002
2802	85745271 (5803, 5804)	Novel Protein sim. GBank gi 2414615 emb CAB16364  - (Z99259) hypothetical protein [Schizosaccharomyces pombe]				264683, 264691
2803	87606733 (5805, 5806)	Novel Protein sim. GBank gi 1079318 pir J552241 - XLCL2 protein - African clawed frog				264887, 22278994, 264259, 29331826, 29331828, 264905, 52644045, 56182435, 264511, 265017, 265018, 18108351, 264448, 264683, 264769, 264689, 35695917
2804	86458072 (5807, 5808)	Novel Protein sim. GBank gi 5639823 gb AAD45885.1 AF14367 - (AF143676) multispanning nuclear envelope membrane protein nurim [Homo sapiens]			UNCLASSIFIED	52644150, 87168518, 60432113, 22279002, 52646385, 22278999, 264259, 35696052, 265011, 265017, 264683, 264769, 35695917, 265020, 263967, 18108374, 35695855, 264637, 264952, 18108385, 18108387
2805	84449926 (5809, 5810)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - !!!!! ALU SUBFAMILY SQ WARNING ENTRY !!!!			oncogene	265009, 264681, 264682
2806	95341051 (5811, 5812)	Novel Protein sim. GBank gi 4689256 gb AAD27831.1 AF12185 - (AF121858) sorting nexin 8 [Homo sapiens]	Contains protein domain (PF00787) - PX domain		UNCLASSIFIED	22278996, 35696286, 22278998, 264259, 60432289, 29331828, 29331830, 66712502, 265009, 60170831, 33108954, 264448, 264683, 264288, 264689, 21906766, 21906767, 21906768, 55811957, 35695917, 265022, 52644150, 264691, 33657023, 264692, 264693, 35695855, 60432113, 264566

2907	91211303 (5813, 5814)	Novel Protein sim. GBank gi 1707079 (U80451) - contains strong similarity to a DNAJ-like domain (PS:PS00536) [Caenorhabditis elegans]	Contains protein domain (PF00226) - DnaJ domain	eph	52844507, 56182575, 56181686, 22278995, 56994075, 35696286, 60432049, 56182181, 35696052, 60431735, 264595, 55812038, 21906754, 55811386, 265019, 264682, 264369, 56181562, 21906766, 55811957, 35695917, 265020, 265021, 33657023, 33657109, 60431528, 55811576, 35696423, 35695855, 264638, 22279000
2908	80414246 (5815, 5816)	Novel Protein sim. GBank gi 2673917 (AC002561) - putative ATP-dependent RNA helicase [Arabidopsis thaliana]		helicase	265009, 33103954, 18108351, 264766, 265021, 284691, 264892, 18108374, 264556, 264638, 264557, 264558
2909	87420225 (5817, 5818)			eph	264259, 87168474, 265018, 18108365, 264628
2910	86601075 (5819, 5820)	Novel Protein sim. GBank gi 4539335[emb]CAB37483.1] - (AL035539) putative protein [Arabidopsis thaliana]			22278995, 264509, 264512, 265007, 33657402, 265017, 264369, 265022, 18108365, 264628
2911	94216615 (5821, 5822)	Novel Protein sim. GBank gi 4469187[emb]CAB38415.1] - (AL031588) dJ1163J1.3 (novel protein similar to mouse B99) [Homo sapiens]	glucoamylase		52646365, 18108397, 22278995, 22278997, 22278998, 22278999, 29331824, 29331825, 52644045, 265008, 265018, 264448, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265021, 18108370, 18108372, 18108374, 22279000
2912	87731803 (5823, 5824)	Novel Protein sim. GBank gi 4928637[gb]AAD34079.1 AF15184 - (AF151842) CGI-84 protein [Homo sapiens]	Contains protein domain (PF00904) - Involucrin repeat		52645156, 264092, 60432049, 264259, 52645080, 29331824, 29331825, 66712502, 33109984, 264760, 264883, 264288, 264686, 265021, 264693, 18108368, 263976, 264404
2913	87713823 (5825, 5826)	Novel Protein sim. GBank gi 854065[emb]CAA58337] - (X83413) U88 [Human herpesvirus 6]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	52644507, 52645156, 52646842, 56182575, 35696286, 22278997, 264259, 52645080, 29331827, 35696052, 29331828, 264828, 52644045, 56182435, 55812038, 52646317, 21906754, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264882, 264686, 264689, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 52644150, 27486261, 27486262, 27486265, 35695763, 55811576, 35695855, 52644332, 22279000, 22279002, 264557
2914	87797300 (5827, 5828)				264557

2915	88081972 (5829, 5830)	Novel Protein sim. GBank gij5174485[ref]NP_008030.1[p]KIAA - endocytic receptor (macrophage mannose receptor family)	Contains protein domain Lectin C-type domain	- eph	264569, 264488, 264687, 264768, 21808760, 52646842, 21906767, 21806768, 56182575, 29148629, 35695917, 22278996, 22278997, 22278998, 265021, 22278999, 52644150, 264691, 264259, 60432049, 264692, 52645129, 33657109, 33657182, 29331827, 27466261, 35696052, 29331828, 27486262, 27486264, 27486265, 33657349, 29148498, 29146499, 264906, 264907, 18108370, 264908, 18108372, 52644045, 18108374, 56182435, 35695855, 264112, 264510, 265008, 60432229, 264593, 60433356, 56182323, 18108382, 55812038, 18108385, 33109954, 21906754, 33657084, 87168518, 265010, 265011, 60432113, 265017, 265018, 22279000, 265019, 55811150, 264681, 18108351, 264763, 264448, 264683, 264566, 18108354, 264389, 264288, 264766, 52645156, 65274572, 22278994, 22278995, 35696288, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 33656970, 264509, 264906, 29331830, 52644045, 264909, 56182435, 60170831, 264592, 264593, 33657402, 60433356, 52646317, 21908754, 33109954, 33657084, 52644296, 85658542, 265011, 265017, 265018, 265019, 18108351, 264448, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265021, 265022, 52644150, 33657023, 33657109, 33657182, 27486261, 27486262, 27486264, 35695763, 18108376, 55811576, 35696423, 65274791, 35695855, 52644332, 264557, 264638, 56182323, 18108387, 87168518, 22279002, 264482
2916	95337780 (5831, 5832)	Novel Protein sim. GBank gij5104851[dbj]BAA80165.1) - (AP000061) 305aa long hypothetical dTDP-4- dehydrohamose reductase [Aeropyrum pernix]		dehydrogenase	
2917	87454546 (5833, 5834)	Novel Protein sim. GBank gij3169065[emb]CAA19260.1) - (AL023704) putative translocation elongation factor-Tu fa mily [Schizosaccharomyces pombe]		UNCLASSIFIED	60433438, 264602, 264682, 87168518, 60432113
2918	85690529 (5835, 5836)	Novel Protein sim. GBank gij539218[pir]S38038 - hypothetical protein YKL201c - yeast (Saccharomyces cerevisiae)			264638
2919	87641497 (5837, 5838)	Novel Protein sim. GBank gij2564855 (AF030001) - unknown [Mus musculus]			66714117, 66712502, 263981

2920	87769523 (5839, 5840)				3566286, 2227897, 264259, 52645080, 29331824, 29331826, 29331827, 264828, 264909, 56182435, 264511, 264758, 33109954, 21908754, 52644296, 265010, 285011, 264601, 265017, 285019, 284681, 264687, 21908767, 265021, 52644150, 264690, 264691, 264692, 264693, 33657109, 33657182, 27486262, 27486264, 27486265, 35696423, 35695855, 264632, 264636, 264637, 264638, 56182323, 60170394, 18108385, 87168518, 60432113
2921	91639982 (5841, 5842)	Novel Protein sim. GBank gi 4580013 gb AAD24202.1 U83194_ - (U83194) TRAF4-associated factor 2 [Homo sapiens]	Contains protein domain (PF00787) - PX domain		35696286, 2227897, 264091, 264092, 264094, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 29146498, 264104, 264105, 264107, 264509, 264110, 264112, 264512, 60433356, 21906754, 87168474, 265017, 18108351, 264288, 21906765, 21908766, 21906767, 21906769, 35695917, 265021, 263974, 18108374, 263976, 263977, 18108376, 264555, 263981, 56528486, 87168518, 22279000, 22279002
2922	87749762 (5843, 5844)	Novel Protein sim. GBank gi 4589514 dbj BAA76779.1  - (AB023152) KIAA0935 protein [Homo sapiens]	Contains protein domain (PF01074) - Glycosyl hydrolases family 38	kinase	264908, 264909, 264511, 265006, 265008, 264593, 33657402, 60174639, 18108351, 264763, 21906765, 29148627, 35695917, 264692, 264629, 263978, 55811576, 35695855, 264555, 264558, 56182323, 60170394, 22279000, 264486
2923	95337789 (5845, 5846)	Novel Protein sim. GBank gi 4835268 emb CAB42898.2  - (Z83844) dJ37E16.4 (similar to mouse p116Rip protein) [Homo sapiens]	Contains protein domain (PF00169) - PH domain	struct	264488, 18108397, 22278995, 22278996, 22278997, 22278998, 22278999, 29331825, 29331826, 29331827, 29331830, 264511, 265009, 33657402, 265011, 265017, 265018, 264883, 18108354, 21906765, 21906767, 21906768, 21906769, 52644150, 264691, 264692, 33657109, 263974, 18108376, 264631, 264636, 18108385, 18108387, 22279000, 264563, 264566
2924	87791967 (5847, 5848)	Novel Protein sim. GBank gi 2133095 pir S72254 - ribosomal protein L36, mitochondrial - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00444) - Ribosomal protein L36	ribosomalprot	265017, 264628, 20281152, 264556
2925	95090120 (5849, 5850)	Novel Protein sim. GBank gi 2388986 emb CAB11718  - (Z88880) actin associated protein [Schizosaccharomyces pombe]		UNCLASSIFIED	56182575, 35686286, 264259, 60432289, 29331827, 264508, 52644045, 264910, 264591, 60432229, 55812038, 21906754, 264681, 264448, 264683, 264288, 264685, 52644229, 264689, 21906765, 21906766, 21908768, 21908769, 265021, 265022, 60170615, 264692, 33657023, 264693, 33657109, 35696423, 65274791, 56182323

2928	95343003 (5851, 5852)	Novel Protein sim. GBank gi 283032 pir S22456 - hydroxyproline-rich glycoprotein - perennial teasinte			29331828, 265011, 264768, 264689
2927	80408018 (5853, 5854)				264764, 264288, 264630, 264637
2928	20452179 (5855, 5856)	Novel Protein sim. GBank gi 3413320 emb CAA06915  - (AJ006215) CMP-N-acetylneuraminic acid synthetase [Mus musculus]	UNCLASSIFIED		264558
2929	91622920 (5857, 5858)		UNCLASSIFIED		264569, 264489, 22278994, 35696286, 22278996, 22278998, 22278999, 264094, 264259, 52645080, 29331822, 29331824, 66714117, 29331825, 29331826, 29331827, 35696052, 33856970, 264109, 29331830, 52644045, 265009, 33109954, 52644296, 87168559, 264760, 264762, 264448, 264764, 264288, 264766, 264768, 21806765, 21908768, 21906768, 21906769, 35695917, 264691, 33657023, 264693, 33657109, 18108374, 263976, 35698423, 35695855, 263981, 22279000, 22279002, 264567, 264486
2930	95302755 (5859, 5860)		UNCLASSIFIED		56182575, 56181686, 35696286, 22278996, 22278998, 22278999, 264259, 29331825, 60432289, 29331828, 264905, 52644045, 56182435, 265009, 60170831, 264592, 60432229, 60433356, 87168474, 265010, 265011, 265017, 265018, 265019, 264762, 264448, 264683, 264288, 264768, 21906765, 21906769, 35695917, 60170815, 33657023, 33657109, 264628, 18108370, 18108372, 35696423, 35695855, 264556, 56182323, 60432113, 264567
2931	94312683 (5861, 5862)	Novel Protein sim. GBank gi 3786433 (AF098505) - similar to Arabidopsis thaliana male sterility protein 2 (SW/Q08891) [Caenorhabditis elegans]	UNCLASSIFIED	Contains protein domain (PF00471) - Ribosomal protein L33	52645156, 22278997, 22278998, 29331822, 52645080, 29331824, 60432289, 33656970, 60433356, 60433438, 33109954, 21806765, 21906766, 21906767, 21906768, 265020, 52644150, 33657023, 33657109, 33657182, 27486265, 35698423, 35695855, 264555, 87168518, 60432113, 264566
2932	79632623 (5863, 5864)				264906, 264907
2933	91720776 (5865, 5866)	Novel Protein sim. GBank gi 3378056 (AF017777) - helicase [Drosophila melanogaster]	helicase		264488, 18108392, 56182575, 22278999, 264091, 264259, 29331825, 60432289, 29331827, 264508, 52644045, 56182435, 265007, 265009, 264592, 60433356, 60433438, 21906754, 265017, 264682, 264288, 52644229, 21906765, 21906766, 21908768, 21906769, 265022, 52644150, 33657023, 33657109, 27486265, 264635, 264636, 60170394, 56182323, 18108385, 60432113, 264565, 264566, 264567

2934	86576025 (5867, 5868)				22278997, 22278999, 29331824, 33657402, 264691, 27486262, 264628, 87168518, 22279000
2935	86410579 (5869, 5870)			UNCLASSIFIED	56182575, 22278995, 60433356, 33657402, 264758, 33109954, 21906754, 265018, 265019, 264448, 264769, 21906764, 21906765, 265021, 264692, 33657023, 33657109, 33657349, 55810764, 22279000
2936	87605863 (5871, 5872)	Novel Protein sim. GBank gi 4153862 (AC005065) - determined by GENSCAN prediction and spliced EST; match to EST R84329 (NID:942735) [Homo sapiens]	Contains protein domain (PF00856) - SET domain	nuclease	22278997, 29331827, 29331828, 265009, 265017, 264605, 265020, 55811576, 18108387, 60432113, 264563
2937	94853096 (5873, 5874)	Novel Protein sim. GBank gi 5174409 [NP_006101.1] pCD28 - CD2 antigen (cytoplasmic tail)-binding protein 2		UNCLASSIFIED	56994075, 22278999, 264259, 60432049, 29331822, 56182181, 29331827, 29331828, 264906, 264908, 264909, 56182435, 265006, 264512, 264910, 60170831, 60433356, 265011, 265018, 18108351, 264448, 264288, 264766, 52644228, 21908765, 29148784, 65274791, 264556, 56182323, 60170394, 264558, 60432113, 264565, 264486, 264567
2938	95419773 (5875, 5876)	Novel Protein sim. GBank gi 3318990 [emb] CAA76720 - (Y17287) ubiquitin-conjugating enzyme [Mus musculus]	Contains protein domain (PF00179) - Ubiquitin-conjugating enzyme	ubiquitin	264488, 56182575, 22278998, 35696286, 22278997, 22278998, 22278999, 264490, 264259, 29331822, 29331824, 66714117, 29331827, 35696052, 264107, 264905, 66712502, 52644045, 56182435, 264511, 265008, 265009, 60432228, 33657402, 60433438, 55812038, 21806754, 85658542, 265010, 265011, 87168559, 265017, 265018, 265019, 264681, 264288, 264689, 21906765, 21906767, 21906768, 55811957, 35695917, 265020, 60170615, 264690, 264691, 264692, 33657023, 264683, 65274620, 33657109, 18108370, 18108374, 263976, 35696423, 35695855, 264555, 264556, 18108381, 56182323, 60170394, 83373044, 18108385, 56526486, 60432113, 22279002
2939	87786622 (5877, 5878)	Novel Protein sim. GBank gi 3979900 [emb] CAA99909 - (Z75547) similar to WD domain, G-beta repeat: cDNA EST yk371b7.5 comes from this gene; cDNA EST yk312h1.5 comes from this gene; cDNA EST yk465d5.5 comes from this gene; cDNA EST yk472c4.5 comes from this gene; cDNA EST yk292f8 ....	Contains protein domain (PF00400) - WD domain, G-beta repeat	ATPase-associated	264907, 265018, 264681, 264685, 264686



2840	95011103 (5879, 5880)			UNCLASSIFIED	22278996, 29331822, 29331824, 66714117, 29331826, 29331828, 264905, 264908, 66712502, 29331830, 265011, 265017, 264764, 264369, 21906766, 21906767, 33657023, 33657109, 32833986, 18108374, 18108377, 264634, 83373044, 18108385, 18108387, 264566
2841	21423370 (5881, 5882)	Novel Protein sim. GBank gi 3413872 dbj BAA32300  - (AB007924) KIAA0455 protein [Homo sapiens]		UNCLASSIFIED	264557
2842	87430203 (5883, 5884)	Novel Protein sim. GBank gi 1172845 sp P46629 RB25_RABIT - RAS-RELATED PROTEIN RAB-25		glycoprotein	264910, 265010, 264768
2843	95314504 (5885, 5886)	Novel Protein sim. GBank gi 4929653 gb AAD34087.1 AF15185 - (AF15185) CGI-92 protein [Homo sapiens]		collagen	60432049, 264259, 60432289, 29331827, 29146498, 265008, 264593, 60433356, 60433438, 265010, 265011, 265017, 265018, 264683, 264766, 18108381, 65274727, 60432113, 264567
2844	95081063 (5887, 5888)	Novel Protein sim. GBank gi 4678282 emb CAB41190.1  - (AL049860) 1-acylcerol-3-phosphate acyltransferase-like protein [Arabidopsis thaliana]	Contains protein domain (PF00415) - ATPase-associated Regulator of chromosome condensation (RCC1)		56994075, 22278998, 60432049, 264259, 29331822, 29331824, 60424269, 60432289, 29331826, 29331828, 264905, 264907, 52644045, 264909, 264511, 265006, 265009, 264594, 21906754, 87168559, 264603, 265017, 265018, 18108351, 264682, 264766, 264687, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 60170615, 52644150, 264690, 264691, 33657023, 264692, 264693, 33657109, 33657182, 33657349, 18108370, 18108374, 18108377, 55811576, 35696423, 35695855, 264635, 264555, 264556, 58182323, 60170394, 264558, 264559, 83373044, 58526486, 87168518, 60432113, 22279002, 264482, 264563, 264484, 264567
2845	84233560 (5889, 5890)	Novel Protein sim. GBank gi 728831 sp P39188 ALU1_HUMAN - !!!!! ALU SUBFAMILY J WARNING ENTRY !!!!!	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	60424179, 22278995, 22278996, 22278998, 22278999, 264259, 50182181, 29331824, 60424269, 60432289, 35695052, 264908, 265008, 60433356, 55812038, 264759, 55811386, 265018, 264681, 18108351, 264448, 264683, 264369, 264288, 264687, 56181562, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 264693, 60431528, 55810764, 35696423, 35695855, 264630, 60170394, 83373044, 22279000, 264566, 264567

2846	84317315 (5891, 5892)	Novel Protein sim. GBank gi 5441952 gb AAD43195.1 AF072864 - (AF072864) peroxisomal membrane protein PMP 24 [Homo sapiens]		UNCLASSIFIED	264488, 264259, 264508, 264509, 264906, 264907, 264809, 264510, 264511, 265007, 264512, 264910, 264591, 264593, 18108351, 264764, 264288, 264684, 264769, 265021, 264692, 33657109, 264628, 264629, 18108374, 264631, 264634, 264636, 264637, 18108380, 264638, 264639, 83373044, 264555, 264586, 264488, 264567
2847	87362852 (5893, 5894)	Novel Protein sim. GBank gi 540291 gb AAC34383.1  - (AF056116) Ali-1 related protein [Fugu rubripes]		UNCLASSIFIED	22278995, 22278996, 22278997, 22278998, 29146498, 264508, 29331830, 265007, 265008, 285009, 60432228, 21908754, 265010, 265017, 265019, 264766, 264685, 21908765, 21908766, 21908767, 21908768, 21908769, 265020, 264628, 18108370, 264628, 264630, 18108387, 60432113
2848	87626527 (5895, 5896)	Novel Protein sim. GBank gi 5566614 gb AA865654.2  - (AF001533) mitogen-induced [Mus musculus]			52646842, 22278995, 264259, 29331824, 29331825, 29331827, 29331830, 264909, 265007, 265009, 265019, 264763, 264684, 264288, 264685, 264888, 21908767, 264691, 264692, 264693, 18108374, 55811576, 18108385, 22279002, 264553, 264567
2849	88175545 (5897, 5898)	Novel Protein sim. GBank gi 2132923 pr S67133 - probable membrane protein YOR240w - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED	22278996, 22278997, 60432289, 29331826, 29331827, 29331828, 35696052, 29146499, 264104, 264107, 264905, 66712502, 264908, 60433356, 60433438, 87168559, 264764, 52644229, 56181562, 21908767, 21908768, 21908769, 265022, 60170615, 33657023, 35696423, 263981, 264558, 60432113, 22279002
2850	85086870 (5899, 5900)	Novel Protein sim. GBank gi 466102 sp P34629 YOJ6_CAEEL - PUTATIVE AMINOPEPTIDASE ZK353.6 IN CHROMOSOME III	Contains protein domain (PF00883) - Cytosol aminopeptidase family	peptidase	264488, 35696286, 264259, 35696052, 264907, 265007, 264910, 265017, 265018, 264288, 264768, 35695917, 265020, 18108362, 18108370, 18108379, 35696423, 65274791, 35695855, 264556, 56526486, 264486
2851	87392357 (5901, 5902)	Novel Protein sim. GBank gi 468902 emb CAB41450.1  - (AJ238248) centaurin beta2 [Homo sapiens]			264693

2952	95329952 (5903, 5904)	Novel Protein sim. GBank gij559693jembj(CAB51405.1) - (AL096861) hypothetical protein [Homo sapiens]	Contains protein domain (PF00650) - CRAL/TRIO domain.	transcript factor	264687, 52645156, 21908768, 21908769, 22278996, 265020, 264690, 60432049, 264259, 264693, 29331822, 18108365, 29331825, 60432289, 33657109, 18108368, 29331827, 35696052, 27486262, 264508, 264905, 20281149, 264906, 264907, 29331830, 264908, 264909, 35695855, 264511, 265008, 265009, 264910, 264635, 264636, 60432229, 264638, 60433358, 264639, 264758, 87188518, 265017, 22279000, 22279002, 264760, 264563, 264482, 18108351, 264448, 264288
2953	88093575 (5905, 5906)	Novel Protein sim. GBank gij119522jsp10658jSERC_RABIT - PROBABLE PHOSPHOSERINE AMINOTRANSFERASE (PSAT1) (ENDOMETRIAL PROGESTERONE-INDUCED PROTEIN) (EPIP)	Contains protein domain (PF00266) - Amino transferases class-V	UNCLASSIFIED	18108396, 56994075, 22278996, 29331822, 29331824, 29331825, 29331827, 35896052, 264508, 264905, 264908, 264907, 264510, 264591, 264594, 33657402, 264595, 264596, 264758, 52646317, 21908754, 33657084, 52644296, 87188559, 264600, 264760, 264681, 18108351, 264764, 264369, 264288, 264687, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 33657023, 18108364, 52645128, 33657109, 33657349, 18108374, 263978, 35695855, 264637, 264638, 87188518, 264482, 264563, 264565
2954	88086288 (5907, 5908)	Novel Protein sim. GBank gij4885261jrefjNP_005251.1pGDF9 - growth differentiation factor 9	Contains protein domain (PF00019) - Tgf transforming growth factor beta like domain	tgf	29331822
2955	87698426 (5909, 5910)	Novel Protein sim. GBank gij3452473 (AF084205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264259, 29331822, 29331824, 29331825, 29331826, 35696052, 264908, 52644045, 264512, 60432229, 265018, 265019, 55811150, 264769, 21906767, 21906768, 21908769, 265021, 60170615, 55810764, 264567
2956	85789745 (5911, 5912)	Novel Protein sim. GBank gij4689254jgbjAAD27830.1jAF12185 - (AF121857) sorting nexin 7 [Homo sapiens]	Contains protein domain (PF00787) - PX domain		22278996, 264259, 29331827, 264808, 21906768
2957	80933301 (5913, 5914)	Novel Protein sim. GBank gij4503023jrefjNP_000089.1pCPT2 - carnitine palmitoyltransferase II precursor		cadherin	22278999, 264259, 29331824, 29331827, 285008, 264595, 264758, 265010, 265011, 264448, 264763, 264683, 264288, 264685, 18108357, 29148629, 264690, 18108362, 264693, 18108370, 60431528, 18108374, 264634, 18108381, 56182323, 18108382, 18108385, 18108388, 56526486, 87188518, 264482, 264487
2958	87440014 (5915, 5916)	Novel Protein sim. GBank gij4240257jdbjBAA74907.1 - (AB020691) KIAA0884 protein [Homo sapiens]			264595, 264596, 264681, 264368, 264629, 264631, 264567

2959	95109420 (5917, 5918)	Novel Protein sim. GBank gj988221 (U33005) - Tbc1 [Mus musculus]	Contains protein domain (PF00566) - TBC domain	oncogene	263994, 22278997, 264259, 60432049, 29331826, 29331828, 35896052, 29331830, 66712502, 56182435, 265006, 264512, 265008, 265009, 60433356, 60433438, 264596, 265017, 265018, 264683, 264288, 264766, 264769, 21906766, 21906767, 21906769, 265020, 60170615, 264692, 27486265, 18108374, 65274791, 35695855, 83373044, 56526486, 60432113
2960	87420091 (5918, 5920)			UNCLASSIFIED	35696286, 56182435, 87168474, 265010, 60170615, 35696423, 56182323, 18108383, 87168518, 264483
2981	95413418 (5921, 5922)	Novel Protein sim. GBank gj596846[embj(CAB05177.2)] - (Z82266) predicted using GeneFinder; similar to WD domain, G-beta repeats [Caenorhabditis elegans]	Contains protein domain (PF00400) - WD domain, G-beta repeat	transcriptfactor	22278997, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 264907, 264908, 52644045, 265006, 33657402, 21906754, 87168474, 265011, 87168559, 265017, 21906769, 265020, 60170615, 264692, 33657023, 35695763, 18108370, 18108374, 35696423, 264632, 264636, 18108385, 87168518, 22279002, 264564, 264567
2982	87912700 (5923, 5924)			UNCLASSIFIED	35696286, 22278997, 264092, 264094, 264259, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264508, 264905, 264509, 264907, 264908, 264909, 264510, 264512, 264593, 264594, 60433438, 264758, 52646317, 264802, 264603, 264605, 264760, 264762, 264764, 264288, 264766, 264686, 264768, 264769, 35695917, 265020, 264691, 264634, 264636, 264637, 264638, 264639, 18108385, 264563, 264565, 264566, 264567, 264486
2983	95313464 (5925, 5926)	Novel Protein sim. GBank gj4240223[dbj BAA74890.1] - (AB020674) KIAA0867 protein [Homo sapiens]	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	transcriptfactor	18108392, 56994075, 22278998, 22278999, 29331822, 29331825, 29331826, 29331827, 29331828, 265007, 265008, 264592, 264594, 21906754, 265018, 264760, 264687, 29148627, 29148784, 265020, 33657023, 264693, 65274620, 33657182, 27486281, 264629, 55810764, 35696423, 264555, 264636, 264637, 264557, 264558, 264563
2984	94324617 (5927, 5928)			UNCLASSIFIED	264259, 29331828, 33657402, 265017, 265018, 264692, 18108388, 83373044, 18108388

2865	80384762 (5929, 5930)	Novel Protein sim. GBank gi 485447 ref NP_005452.1 pKRM1 - Kreiser (mouse) maf- related leucine zipper homolog			transcriptfactor	264259, 26331826, 264508, 264509, 264905, 264907, 264908, 264909, 264511, 265008, 264910, 264591, 264593, 264594, 33657402, 265011, 264760, 264762, 264764, 264288, 264685, 264766, 264692, 33657109, 264628, 264629, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 264563, 264567, 18108391 60432289, 264682, 264448
2866	91725248 (5931, 5932)	Novel Protein sim. GBank gi 5262751 emb CAB45690.1  - (AJ243177) Xenopus RPA interacting protein alpha [Xenopus laevis]				
2867	84658303 (5933, 5934)	Novel Protein sim. GBank gi 524225 (U19181) - Rabin3 [Rattus norvegicus]			UNCLASSIFIED	264488, 264508, 264509, 264908, 264909, 264511, 264910, 264594, 264758, 85658542, 264762, 264764, 265021, 264556, 18108381, 264564, 264488
2868	95302776 (5935, 5936)	Novel Protein sim. GBank gi 4929715 gb AAD34118.1 AF15188 - (AF151881) CGI-123 protein [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)			264687, 52645156, 21906765, 52646365, 21906767, 18108398, 35696423, 22278996, 35696286, 22278997, 265020, 22278999, 265021, 265022, 264093, 264638, 264690, 52644150, 264259, 33657023, 52645080, 264693, 29331822, 56182181, 29331824, 66714117, 29331825, 33109954, 52645129, 29331826, 21906754, 33657182, 29331827, 29331828, 35696052, 27486262, 87168518, 87168474, 265010, 87168559, 265018, 22278000, 265019, 22279002, 264563, 18108351, 264906, 264907, 264448, 66712502, 264566, 264369, 264288
2869	95310957 (5937, 5938)	Novel Protein sim. GBank gi 3024743 sp Q24734 THSA_SULS7 - THERMOSOME, ALPHA SUBUNIT (CHAPERONIN ALPHA SUBUNIT)		eph		52646842, 22278996, 22278998, 22278999, 60432049, 264259, 29331824, 29331825, 29331826, 29331828, 264509, 264909, 52644045, 56182435, 265009, 60433438, 55812038, 21906754, 265011, 87168559, 265018, 265019, 264448, 264288, 264369, 52644228, 21906766, 21906768, 21906769, 29148784, 265020, 265021, 52644150, 264691, 33657109, 18108374, 56182323, 60170394, 87168518, 60432113, 22279000
2870	88088071 (5939, 5940)	Novel Protein sim. GBank gi 3165407 (AC004755) - fos37502_1 [Homo sapiens]	Contains protein domain (PF00046) - Homeobox domain	homeobox		

2971	94196930 (5941, 5942)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		Im7	264488, 56182575, 35586286, 56994075, 29331824, 29331826, 29146499, 264508, 264905, 264907, 264112, 264910, 21906754, 87168559, 265018, 265019, 18108351, 264689, 21906765, 21906767, 21906768, 265020, 265021, 60170815, 18108364, 264628, 264629, 18108374, 264636, 264556, 264558, 83373044, 18108384, 18108385, 87168518, 264564, 264567
2972	86625943 (5943, 5944)	Novel Protein sim. GBank gi 728836 sp P39193 ALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII		kinase	265017, 35695917, 265021, 33657109, 22279002, 264583
2973	91215301 (5945, 5946)	Novel Protein sim. GBank gi 2746789 (AF040842) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	29331822, 264692, 33657349, 55811576, 264563
2974	91673002 (5947, 5948)	Novel Protein sim. GBank gi 786117 (L41834) - nuclear protein [Ensis minor]		UNCLASSIFIED	18108392, 52644507, 56182575, 56181686, 22278995, 22278996, 35696288, 22278997, 22278998, 22278999, 264259, 52845080, 29331824, 29331825, 66714117, 60424269, 29331826, 29331827, 29331828, 35696052, 66712502, 264908, 52644045, 265007, 264910, 265009, 60433438, 33109954, 21906754, 55811386, 52644286, 87168474, 87168559, 265017, 265018, 265019, 18108351, 264446, 264369, 264288, 52644229, 18108359, 21906765, 21906767, 21906768, 35695917, 265020, 265021, 52644150, 264691, 264692, 33657023, 27486262, 27486264, 35695763, 18108370, 18108376, 55810764, 55811578, 35696423, 35695955, 264630, 264635, 264557, 52644332, 264558, 83373044, 18108387, 87168518, 60432113, 22279000, 264482, 264487
2975	95325213 (5949, 5950)	Novel Protein sim. GBank gi 3880812 emb CAA19508  - (AL023839) similar to HECT-domain (ubiquitin-transferase), cDNA EST yk480d10.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00632) - HECT-domain (ubiquitin- transferase).	ubiquitin	29331824, 29331827, 29331828, 264910, 85658542, 265011, 265018, 264448, 264288, 264769, 21906767, 265020, 264691, 264559, 83373044
2976	87771202 (5951, 5952)	Novel Protein sim. GBank gi 5678136 gb AAD46874.1 AF16093 - (AF160934) BcDNA.LD14189 [Drosophila melanogaster]		transport	22278996, 264906, 265007, 265010, 265011, 265017, 265019, 18108351, 264685, 264689, 18108370, 264639, 18108385
2977	91725254 (5953, 5954)	Novel Protein sim. GBank gi 5262751 emb CAB45690.1  - (A1243177) Xenopus RPA interacting protein alpha [Xenopus laevis]		UNCLASSIFIED	264509, 264288

2878	87332059 (5955, 5956)	Novel Protein sim. GBank gij746549 (U23522) - No definition line found [Caenorhabditis elegans]	Contains protein domain (PF00480) - ROK family	UNCLASSIFIED	22278995, 22278996, 22278997, 22278999, 264259, 60432289, 29331827, 29146499, 56182435, 265006, 265007, 265009, 60433356, 60433438, 21906754, 265010, 265011, 265017, 265018, 265019, 264288, 264685, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 35696423, 264639, 60432113, 22279000, 22279002
2879	91725256 (5957, 5958)	Novel Protein sim. GBank gij5262751[embjCAB45690.1] - (AJ243177) Xenopus RPA Interacting protein alpha [Xenopus laevis]		complement	264488, 65274572, 56994075, 22278999, 264093, 29331822, 29331824, 264288, 55811957, 33657023, 33657109, 18108370, 55811576, 56182323, 60432113, 264482
2880	86296600 (5958, 5960)				265009, 21906767, 263981, 22279000
2881	87376330 (5961, 5962)			UNCLASSIFIED	264629, 264564
2882	95303675 (5963, 5964)	Novel Protein sim. GBank gij4928767[gbjAAD34144.1]AF15190 - (AF151907) CGI-149 protein [Homo sapiens]			22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264092, 29331824, 29331827, 29331828, 264905, 264591, 264592, 264594, 264595, 264596, 33657084, 264448, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265022, 18108365, 33657182, 33657349, 35696423, 83373044, 22279000, 22279002
2883	91725258 (5965, 5966)	Novel Protein sim. GBank gij5262751[embjCAB45690.1] - (AJ243177) Xenopus RPA Interacting protein alpha [Xenopus laevis]			60424178, 52646842, 18108398, 22278997, 264093, 60432049, 264259, 29331822, 60432288, 33656970, 264905, 52644045, 265006, 60431735, 87168474, 265018, 265019, 18108351, 264448, 21906765, 21906768, 35695917, 33657023, 52845129, 18108370, 35696423, 83373044, 56526486, 60432113, 264404, 22279002
2884	94136467 (5967, 5968)	Novel Protein sim. GBank gij2393734 (AC002542) - similar to C. elegans F11A10.5; 80% similarity to Z68297 (PID:g1130618) [Homo sapiens]		ATPase associated	
2885	87099072 (5969, 5970)	Novel Protein sim. GBank gij103160[pirjS22126 - finger protein unkempt - fruit fly (Drosophila melanogaster)]		UNCLASSIFIED	264910, 55812038, 56181562, 55811957, 264628, 55810764, 264632, 264635, 60432113
2886	86284861 (5971, 5972)				55811957, 264566
2887	86455934 (5973, 5974)			UNCLASSIFIED	264369

2988	95357753 (5975, 5976)	Novel Protein sim. GBank gij4679028[gb AAD27002.1] - (AF077207) HSPC021 [Homo sapiens]	UNCLASSIFIED	264488, 65274572, 22278995, 22278996, 22278997, 22278999, 264092, 264094, 264259, 60432049, 29331824, 29331826, 60432289, 35696052, 29331828, 264107, 264905, 264907, 264908, 66712502, 264828, 264909, 56182435, 265006, 265007, 265008, 60170831, 60432229, 264593, 60433356, 264757, 60433438, 21908754, 265010, 265011, 87168559, 265017, 265018, 264682, 264448, 264369, 264288, 264685, 52644229, 21908765, 21906767, 21908769, 35695917, 265021, 265022, 52644150, 264690, 33657023, 65274620, 263967, 33657109, 27486262, 18108370, 18108372, 18108374, 55810764, 65274791, 35695855, 264635, 264636, 264637, 263981, 264638, 56182323, 83373044, 60432113, 22279000, 264563, 264564, 264565, 264566, 264567, 22278996, 22278997, 264905, 264511, 60170831, 264593, 265019, 21906765, 21906767, 21906768, 18108374
2989	91225118 (5977, 5978)	Novel Protein sim. GBank gij113671[sp P23964 ALUF_HUMAN - III] ALU CLASS F WARNING ENTRY III!	kinase	22278996, 22278997, 264905, 264511, 60170831, 264593, 265019, 21906765, 21906767, 21906768, 18108374
2990	87330444 (5979, 5980)	Novel Protein sim. GBank gij2829836[sp P97348 RHOD_MOUSE - RHO-RELATED GTP-BINDING PROTEIN RHOD	oncogene	265007, 264512, 18108351, 264288, 264689, 265020, 264691, 33657023, 33657109
2991	94325361 (5981, 5982)		UNCLASSIFIED	264563
2992	85425164 (5983, 5984)		UNCLASSIFIED	264259, 265018, 264689, 18108385
2993	94325363 (5985, 5986)		UNCLASSIFIED	264488, 29331822, 265017, 264761, 21906769, 65274791, 263981, 264565
2994	94136634 (5987, 5988)	Novel Protein sim. GBank gij2496549[sp Q50658 YU02_MYCTU - HYPOTHETICAL 28.7 KD PROTEIN CY339.02	transport	22278994, 22278995, 56994075, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 265006, 265009, 264910, 33109954, 87168474, 87168559, 265018, 265019, 264448, 264288, 21906766, 21906767, 21906768, 21906769, 265021, 265022, 33657023, 264693, 35695855, 83373044, 18108385, 22279000, 264565, 264566, 264905, 264907, 265019, 18108351, 264683
2995	87591070 (5989, 5990)	Novel Protein sim. GBank gij2734081 (AF000195) - similar to oxysterol-binding proteins [Caenorhabditis elegans]		
2996	91013788 (5991, 5992)	Novel Protein sim. GBank gij2829912 (AC002291) - Similar ATP-dependent RNA Helicase [Arabidopsis thaliana]	helicase	65274572, 35696286, 264259, 29331824, 35696052, 29146499, 264508, 264907, 265007, 265008, 60433438, 18108348, 265017, 264681, 264683, 264288, 264766, 264769, 264689, 35695917, 60170615, 33657023, 264692, 264634, 264555, 18108381, 18108382, 18108388, 264484



2897	87627440 (5983, 5994)	Novel Protein sim. GBank gij4589652[dbj BAA76848.1] - (AB023221) KIAA1004 protein [Homo sapiens]		homeobox	264488, 56182575, 264259, 66714117, 29331826, 35896052, 264508, 264509, 264907, 264908, 265006, 87168474, 265019, 264448, 264682, 264685, 264766, 21908764, 21908766, 21906768, 21906769, 27486261, 18108374, 35896423, 264634, 264635, 264636, 264557, 18108385, 87168518, 52646365, 22278997, 264508, 264908, 18108351, 21906765, 21906767, 18108370, 18108374, 35896423, 264636, 264639
2998	88095381 (5995, 5996)	Novel Protein sim. GBank gij3947589[emb CAA22252] - (AL034384) cDNA EST yk255b9.3 comes from this gene; cDNA EST yk255b9.5 comes from this gene; cDNA EST EMBL: M75923 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	52646365, 22278997, 264508, 264908, 18108351, 21906765, 21906767, 18108370, 18108374, 35896423, 264636, 264639
2998	94847055 (5997, 5998)	Novel Protein sim. GBank gij115408[sp P18835][CC19_CAEEL - CUTICLE COLLAGEN 19		UNCLASSIFIED	56182575, 22278996, 28147620, 29331825, 29146498, 29146499, 264905, 66712502, 265006, 265009, 21906754, 85658542, 18108351, 29148627, 29148629, 60170615, 33657109, 27486262, 18108370, 18108374, 264556, 264557, 284558, 60170394, 18108385, 264563
3000	95098370 (5999, 6000)	Novel Protein sim. GBank gij1163174 (U32575) - similar to yeast Sec6p, Swiss-Prot Accession Number P32844; similar to mammalian B94, Swiss-Prot Accession Number Q03169; Method: conceptual translation supplied by author [Rattus norvegicus]		UNCLASSIFIED	264887, 22278997, 22278999, 264259, 29331822, 29331824, 35896052, 28146498, 264508, 264905, 264906, 264907, 264908, 264909, 284510, 264511, 265006, 265007, 265008, 265009, 264910, 33657402, 264757, 264595, 284596, 284758, 21908754, 265011, 264600, 265017, 265018, 264605, 265019, 264760, 264761, 264762, 264681, 264682, 264764, 264288, 264685, 264766, 264686, 264768, 264769, 21906765, 21906768, 35895917, 285020, 264691, 264692, 33657023, 284693, 33657109, 33657182, 27486261, 264628, 264629, 18108374, 18108376, 35896423, 35895855, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264556, 264638, 264639, 60170394, 83373044, 20798451, 22279002, 264563, 264486, 264567
3001	88078454 (6001, 6002)	Novel Protein sim. GBank gij2078470 (AC002073) - Putative gene. Genscan predictions confirmed by EST splicing; coded for by human cDNAs AA122029 (NID:g1678048), D31562 (NID:g644442), AA158721 (NID:g1733515), R59840 (NID:g830335) and F13082 (NID:g709111) [Homo sapiens]		calhepsin	18108394, 52646842, 56182575, 29331824, 29331825, 29331827, 264910, 33109954, 52644296, 265017, 265019, 264288, 265020, 265021, 52644150, 264692, 35895763, 55810764, 35896423, 56182323, 18108387, 264563, 264564
3002	87718167 (6003, 6004)	Novel Protein sim. GBank gij3599478 (AF085185) - Myosin-1A [Acanthamoeba castellanii]		UNCLASSIFIED	264488, 29331824, 29331825, 29331826, 29331827, 29331828, 264906, 264510, 265009, 21906754, 264682, 264688, 33657023, 264569

3003	86648079 (6005, 6006)	Novel Protein sim. GBank gij1754859 (U30292) - collagen type XIII alpha-1 chain [Mus musculus]	Contains protein domain (PF01381) - collagen triple helix repeat (20 copies)	284512, 264593, 264564, 264567, 264486
3004	88066876 (6007, 6008)	Novel Protein sim. GBank gij2224629[dj]BAA20802] - (AB002342) KIAA0344 [Homo sapiens]		29331830, 21906769, 264691, 33657109, 293972, 18108385
3005	87794843 (6009, 6010)	Novel Protein sim. GBank gij4680659[gj]AAD27719.1AF132944 CGI-10 protein [Homo sapiens]	Contains protein domain (PF01360) - oxygenase Monooxygenase	29331822, 29331824, 29331827, 60433438, 265011, 265019, 21906766, 21906767, 21906768, 265020, 33657023, 33657349, 60170394, 22279002, 264567
3006	87422224 (6011, 6012)	Novel Protein sim. GBank gij3930525 (AF064447) - sex-determination protein homolog Fem1a [Mus musculus]	Contains protein domain (PF00023) - MHC Ank repeat	264259, 29331822, 264512, 21906754, 265018, 264687, 21906765, 264691, 264555, 264556, 264558, 18108385
3007	90936005 (6013, 6014)	Novel Protein sim. GBank gij2565052 (U80738) - CAGH1a [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	52644507, 52645156, 65274572, 264809, 284512, 265018, 264760, 264448, 264765, 264689, 60170615, 18108374, 20281152, 264636, 52644332
3008	80416249 (6015, 6016)			264905, 264593, 264768, 264636
3009	91213387 (6017, 6018)	Novel Protein sim. GBank gij3127183 (AF082389) - kidney-specific protein [Rattus norvegicus]	Contains protein domain (PF00501) - AMP-binding enzyme	52646842, 56182575, 22278995, 22278996, 264259, 29331825, 29331826, 29331827, 29331828, 35698052, 264508, 264509, 264907, 56182435, 264511, 265007, 264512, 265008, 264757, 264758, 55812038, 264759, 33109954, 21906754, 265010, 265011, 264600, 265017, 265018, 265019, 264760, 18108351, 264288, 264369, 21906764, 21906765, 21906767, 55811957, 265020, 265021, 264691, 18108368, 27486262, 20281149, 18108370, 55811576, 264637, 264556, 264557, 18108381, 264558, 56182323, 264559, 18108385, 18108388, 22279002, 264486
3010	95317217 (6019, 6020)	Novel Protein sim. GBank gij4927370[gj]AAD33084.1AF06797 - (AF067972) DNA cytosine methyltransferase 3 alpha [Homo sapiens]	Contains protein domain (PF01923) - Protein of unknown function	264686, 264687, 21906767, 21906769, 55811957, 22278995, 35695917, 22278996, 22278997, 265020, 265021, 60170615, 264692, 33657023, 29331822, 264693, 18108364, 29331824, 33657109, 60432289, 29331827, 27486261, 29331828, 264508, 264909, 55811576, 35695855, 265008, 264558, 60433438, 83373044, 18108387, 65274727, 60432113, 265017, 22279000, 265019, 264564, 264682, 264764
3011	94323597 (6021, 6022)	Novel Protein sim. GBank gij5052319[gj]AAD38501.1AF11883 - (AF11883) citrin; adult-onset type II citrullinemia protein [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	35696032, 56182435, 264758, 21906754, 265018, 264760, 264762, 18108351, 264682, 264448, 21906766, 65274620, 18108374, 264482, 264584
3012	87753087 (6023, 6024)		UNCLASSIFIED	263972

3013	91238789 (6025, 6026)	Novel Protein sim. GBank gij3702286 (AC005787) - R33374_1 [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	transcriptfactor	264488, 263994, 35696286, 22278997, 264259, 29331824, 60424269, 66714117, 35696052, 264905, 264906, 264907, 264908, 264909, 56182435, 264511, 264512, 264810, 264591, 264592, 264593, 264594, 33657402, 60433438, 264595, 264596, 55812038, 264758, 33109954, 21906754, 265010, 265018, 264604, 264760, 264682, 264683, 264764, 264389, 264288, 264765, 264766, 264686, 264768, 264687, 21906767, 35695917, 265020, 33657023, 264692, 264693, 33657109, 264628, 264629, 55811578, 35698423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 83373044, 264563, 264565, 264566, 264567
3014	79877263 (6027, 6028)	Novel Protein sim. GBank gij3878374[embjCAA93081] - (Z68879) Similarity to Yeast Chi12p protein (PIR Acc. No. S54453); cDNA EST EMBL:D27950 comes from this gene; cDNA EST EMBL:D27949 comes from this gene; cDNA EST EMBL:D33447 comes from this gene; cDNA EST EMBL:D33316 comes from...		ATPase_associated	264760
3015	86995466 (6029, 6030)				22278995, 22278996, 22278997, 264259, 29331824, 29331828, 264806, 265007, 265008, 264910, 265011, 265017, 265019, 264691, 33657109, 18108370, 35695855, 264556, 264564
3016	87759945 (6031, 6032)	Novel Protein sim. GBank gij1168819[spP41733]CC91_YEAST - CELL DIVISION CONTROL PROTEIN 91		UNCLASSIFIED	52644507, 52646842, 56994075, 52645080, 29331822, 29331824, 35696052, 33656970, 52644045, 264596, 33657084, 265017, 265019, 52644229, 21906767, 35695917, 52644150, 33657023, 33657109, 27486261, 27486262, 27486264, 33657349, 27486265, 35695763, 35695855, 87168518
3017	85011154 (6033, 6034)	Novel Protein sim. GBank gij4589658[dbjBAA76851.1] - (AB023224) KIAA1007 protein [Homo sapiens]			264488, 18108397, 22278996, 35696286, 22278999, 264259, 29331822, 60432289, 264908, 29331830, 264909, 56182435, 265006, 265007, 265008, 265009, 264591, 60433356, 60433438, 52646317, 21906754, 55811386, 265010, 265011, 87168559, 265017, 265018, 265019, 264288, 264687, 21906765, 21906766, 21906767, 21906769, 265020, 265022, 65274820, 52645129, 33657109, 33657182, 18108370, 263972, 18108374, 264631, 52644332, 83373044, 18108385, 18108388, 56526486, 87168518, 264404, 60432113, 22279000, 264567

3018	11073891 (6035, 6036)	Novel Protein sim. GBank gj3219332 (AC004020) - Unknown gene product [Homo sapiens]		oncogene	264558 264569, 52644507, 18108394, 65274572, 56182575, 22278994, 22278995, 56994075, 22278998, 22278999, 264259, 29331822, 29331824, 60432289, 29331827, 264908, 56182435, 265007, 265009, 60432229, 264593, 60433356, 55812038, 21906754, 87168474, 265011, 87168559, 265017, 265018, 265019, 264681, 18108351, 264448, 264682, 264683, 18108354, 264685, 264687, 264689, 21906766, 21906768, 21906769, 52644150, 264690, 264691, 33657023, 264692, 264693, 33657109, 52645129, 33657349, 264629, 65274791, 264634, 52644332, 56182323, 18108385, 87168518, 22279000, 22279002, 264563
3020	94318251 (6039, 6040)	Novel Protein sim. GBank gj3414809 (AF061529) - rjs [Mus musculus]	Contains protein domain (PF00415) - ATPase associated Regulator of chromosome condensation (RCC1)		264488, 263994, 35696286, 264259, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264910, 60174639, 264600, 264603, 264760, 264762, 264682, 264763, 264764, 264288, 264369, 264766, 264687, 264688, 264769, 55811957, 35695917, 33657023, 264628, 35696423, 35695855, 264630, 264632, 264634, 264635, 264636, 264637, 264556, 264557, 264638, 264639, 83373044, 18108385, 264564, 264567, 264486 264769, 264629, 264482
3021	80478512 (6041, 6042)	Novel Protein sim. GBank gj3880889[emb]CAB09005] - (Z95559) cDNA EST yk236d4.5 comes from this gene; cDNA EST EMBL:C13455 comes from this gene; cDNA EST yk329g6.5 comes from this gene; cDNA EST CEMSH45R comes from this gene [Caenorhabditis elegans]			
3022	87718500 (6043, 6044)			UNCLASSIFIED	264259, 29331826, 29331828, 264288, 264566
3023	95305484 (6045, 6046)	Novel Protein sim. GBank gj416592[sp]P32323JAGA1_YEAST - A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR	Contains protein domain (PF00614) - Phospholipase D. Active site motif	UNCLASSIFIED	264488, 22278995, 35696286, 22278997, 29331826, 35696052, 264907, 29331830, 52644045, 56182435, 60432229, 264592, 60433356, 60433438, 264689, 21906767, 55811957, 35695917, 265021, 18108376, 263978, 264635, 264558, 22279000
3024	86675305 (6047, 6048)			UNCLASSIFIED	60432049, 264760, 21906769, 55811957, 35695917, 264690, 264555, 264559 264593, 55811576
3025	65706629 (6049, 6050)	Novel Protein sim. GBank gj295671 (L11275) - selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III [Saccharomyces cerevisiae]			

3026	87643662 (6051, 6052)	Novel Protein sim. GBank gjl3024052[sp]P97924[KARI_RAT - KALIRIN (PAM COOH-TERMINAL INTERACTOR PROTEIN 10) (P-CIP10)]		UNCLASSIFIED	22278996, 22278997, 264490, 29331825, 264111, 265007, 60170831, 265010, 87188559, 265019, 21906765, 29148627, 263967, 20281149, 20281069, 263975, 263977, 20281071, 56526486, 22279000
3027	94844563 (6053, 6054)	Novel Protein sim. GBank gjl4929647[gb]AAD34084.1[AF15184 - (AF151847) CGI-89 protein [Homo sapiens]]	Contains protein domain (PF01529) - DHHC zinc finger domain	UNCLASSIFIED	18108394, 22278995, 22278996, 35696286, 22278997, 22278998, 60432049, 264259, 29331822, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264905, 264907, 29331830, 264908, 264909, 264510, 265007, 265008, 265009, 264910, 33657402, 264596, 21906754, 265010, 265011, 87168559, 264600, 265018, 18108351, 264682, 264683, 264764, 264288, 264685, 264687, 264789, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 29148629, 35695917, 265020, 265021, 265022, 52644150, 264692, 33657023, 264693, 52645129, 33657109, 27486281, 18108374, 55811576, 35696423, 65274791, 264636, 264556, 264557, 264638, 60170394, 264639, 264558, 83373044, 18108385, 56526486, 22279000, 22279002
3028	94231997 (6055, 6056)	Novel Protein sim. GBank gjl3080521[emb]CAA18650[] - (AL022599) hypothetical protein [Schizosaccharomyces pombe]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	22278995, 22278996, 22278997, 22278999, 264259, 29331824, 29331827, 35696052, 264908, 265007, 265008, 265009, 60170831, 21906754, 265011, 87168559, 265018, 264762, 264683, 264765, 264689, 21906765, 21906768, 21906769, 29148629, 35695917, 265021, 265022, 33657109, 27486265, 264628, 264629, 18108374, 35696423, 35695655, 264638, 60170394, 22279000, 22279002, 264482, 264564
3029	87619284 (6057, 6058)			UNCLASSIFIED	22278997, 22278999, 29331827, 264905, 264509, 264909, 264510, 264511, 264512, 87168474, 265019, 18108351, 21906768, 264534, 264690, 264693, 263969, 18108370, 264556, 22279000, 22279002, 264482
3030	87544928 (6059, 6060)	Novel Protein sim. GBank gjl3757728[emb]CAA18782[] - (AL022727) dJ80119.1 (olfactory receptor-like protein [hs6M1-1]) [Homo sapiens]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	- tm7	
3031	91677953 (6061, 6062)	Novel Protein sim. GBank gjl4530587[gb]AAD22105.1[] - (AF132000) TADA1 protein [Homo sapiens]		UNCLASSIFIED	22278995, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 35696052, 264908, 265018, 21906765, 21906766, 21906767, 21906768, 265021, 263974, 18108374, 264556, 56526486, 22279000, 22279002

3032	94130124 (6063, 6064)	Novel Protein sim. GBank gi 1019551 (U37429) - similar to M. musculus MER5 and other AHPCTSA proteins (Caenorhabditis elegans)	Contains protein domain (PF00534) - Glycosyl transferases group 1	synthase	22278996, 35696286, 264259, 29331824, 29331828, 264907, 29331830, 264758, 33109954, 87168474, 87168559, 265019, 264288, 21906769, 265021, 264693, 35698423, 35695855, 264636, 56182323, 83373044, 87168518
3033	95308321 (6065, 6066)	Novel Protein sim. GBank gi 5031573 (ref NP_005712.1 pACTR - ARP3 (actin-related protein 3, yeast) homolog	Contains protein domain (PF00022) - Actin	struct	35696286, 264259, 29331826, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 265008, 264591, 21906754, 265010, 265019, 264681, 264369, 21908784, 21908768, 35695917, 33657023, 264628, 35695855, 264632, 264635, 264639, 264482, 264563
3034	80415373 (6067, 6068)			UNCLASSIFIED	264906, 264907, 264510, 264592, 265010, 264762, 264768, 264837, 264638, 264486
3035	91220692 (6069, 6070)	Novel Protein sim. GBank gi 3738207 (emb CAA21262) - (AL031853) conserved ATP-GTP binding protein (Schizosaccharomyces pombe)		UNCLASSIFIED	264636
3036	91718323 (6071, 6072)	Novel Protein sim. GBank. gi 726837 (sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		kinase	264907, 33657402, 265021
3037	95307434 (6073, 6074)	Novel Protein sim. GBank gi 4406590 (gb AAD20040) - (AF131766) Similar to Ena-VASP like protein [Homo sapiens]			265017
3038	95421807 (6075, 6076)	Novel Protein sim. GBank gi 5360093 (gb AAD42865.1 AF15509) NY-REN 18 antigen [Homo sapiens]	Contains protein domain (PF00627) - UBA domain	UNCLASSIFIED	22278996, 22278997, 264259, 264905, 265007, 265009, 60433356, 21906754, 265018, 265019, 18108351, 264687, 21906765, 265020, 265021, 65274620, 27486262, 264636, 56182323, 18108385, 22279000
3039	87332257 (6077, 6078)	Novel Protein sim. GBank gi 4757128 (emb CAB42094.1  - (AJ238717) ZRP protein [Rattus norvegicus]		UNCLASSIFIED	35696286, 29331828, 264109, 264110, 264511, 265007, 21906754, 265011, 264681, 264683, 264687, 21906768, 264691, 18108370, 263972, 264629, 18108374, 263977, 35696423, 264564, 18108391, 264692, 264558, 18108382, 18108385, 264567
3040	80933517 (6079, 6080)	Novel Protein sim. GBank gi 4884278 (emb CAB43247.1  - (AL050037) hypothetical protein [Homo sapiens]			
3041	88312357 (6081, 6082)	Novel Protein sim. GBank gi 3876073 (emb CAB04122.1  - (Z81505) similar to Zinc finger, C3HC4 type (RING finger); cDNA EST EMBL:D28025 comes from this gene; cDNA EST EMBL:D28024 comes from this gene; cDNA EST EMBL:D33210 comes from this gene; cDNA EST EMBL:D33441 comes from this ...		UNCLASSIFIED	56994075, 22278997, 22278998, 29331827, 33656970, 33109954, 21906754, 87168559, 264600, 264683, 21906765, 21908768, 22279002
3042	85749402 (6083, 6084)	Novel Protein sim. GBank gi 790236 (U21156) - sarcolemmal associated protein-2 [Oryctolagus cuniculus]		glycoprotein	264636

3043	87773026 (6085, 6086)	Novel Protein sim. GBank gij454065jemb[CAA58337] - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	35686286, 60424269, 35696052, 264508, 264905, 66712502, 56182435, 55811386, 52644296, 55811150, 35695917, 60170615, 33657109, 18108374, 264634, 60431850
3044	87646182 (6087, 6088)	Novel Protein sim. GBank gij4104922 (AF042276) - o251 homolog [Pseudomonas putida]	Contains protein domain (PF01209) - ubiE/COG5 methyltransferase family	glycoprotein	22278996, 22278998, 22278999, 29331824, 56182435, 264511, 265007, 60170831, 60432229, 60433356, 33109954, 18108351, 264288, 35695917, 18108368, 18108370, 60170394
3045	94127598 (6089, 6090)	Novel Protein sim. GBank gij4589680[dj]BAA76859.1] - (AB023232) KIAA1015 protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_rna_bind	264488, 264259, 35696052, 264508, 264905, 264509, 264906, 264907, 264909, 264511, 265006, 264591, 264593, 33109954, 264604, 264764, 264683, 264288, 264766, 264768, 21906765, 21906768, 55811857, 35695917, 27486262, 18108370, 264628, 18108374, 35695855, 264630, 264632, 264635, 264563, 264564, 264566
3046	88088247 (6091, 6092)			UNCLASSIFIED	22278999, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264906, 52646317, 55811957, 60432113, 22278900, 22279002, 264482, 264564
3047	85089924 (6093, 6094)			UNCLASSIFIED	264488, 22278996, 22278997, 22278999, 29331824, 29331825, 56182435, 264511, 265008, 265009, 265011, 265017, 264766, 21906768, 21906769, 35695917, 52644150, 33657349, 65274791, 35695855, 264555, 60432113, 22279000, 264566
3048	87629419 (6095, 6096)	Novel Protein sim. GBank gij4588034[gb]AAD25962.1]AF09287 - (AF092878) zinc RING finger protein SAG [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	264102, 29148784
3049	88229855 (6097, 6098)	Novel Protein sim. GBank gij5454158[ref]NP_008286.1]pVARS - valyl-HRNA synthetase 1	Contains protein domain (PF01406) - tRNA synthetases class I (C)	UNCLASSIFIED	22278997, 29331826, 264907, 264758, 87168559, 265018, 264448, 21906768, 265020, 33657109, 35695855, 60432113, 22279000
3050	87643678 (6099, 6100)	Novel Protein sim. GBank gij4589642[dj]BAA76843.1] - (AB023216) KIAA0999 protein [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264259, 29331825, 264909, 265007, 264512, 265019, 264288, 21906766, 265020, 264893, 18108385, 56526486, 87168518, 22279002, 264566
3051	87750599 (6101, 6102)				22278997, 264595, 265018, 264288, 264693, 87168518
3052	57108030 (6103, 6104)	Novel Protein sim. GBank gij117528[sp]P14755[CRYL_RABIT - LAMBDA-CRYSTALLIN		dehydrogenase	264534

3053	95350373 (6105, 6106)	Novel Protein sim. GBank gj13947613 emb CAA19465.1 - (AL023828) cDNA EST EMBL:M89008 comes from this gene; cDNA EST yk28243.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	65274572, 56181686, 22278995, 35696286, 22278998, 264259, 60432289, 265008, 265009, 60433438, 21906754, 265010, 87168559, 264603, 265018, 265019, 264763, 264764, 264288, 21906765, 21906766, 21906768, 21906769, 35695917, 18108374, 35696423, 264638, 56182323, 22279000, 264563
3054	86843510 (6107, 6108)	Novel Protein sim. GBank gj1076211 pir S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii		UNCLASSIFIED	35696286, 35696052, 28331830, 264908, 264909, 264512, 284910, 265017, 264804, 264766, 265020, 33657109, 264628, 35695855, 264638, 264564, 264566, 264486
3055	95350537 (6109, 6110)	Novel Protein sim. GBank gj4680855 gb AAD27717.1 AF13294 - (AF132942) CGI-08 protein [Homo sapiens]		transport	60424179, 65274572, 56182575, 35696286, 22278998, 22278999, 60432049, 264259, 60424269, 60432288, 35696052, 56182435, 265006, 265009, 60170831, 60432229, 60431735, 60433356, 264594, 60433438, 21906754, 55811386, 265011, 87168559, 265019, 18108351, 264683, 264288, 264369, 264689, 21906768, 55811957, 35695917, 60170615, 33657023, 65274620, 33657109, 35695763, 60431528, 18108374, 55810764, 55811576, 35696423, 65274791, 264636, 60431850, 18108381, 56182323, 60170394, 18108385, 60432113, 264564, 264565, 264566
3056	91661636 (6111, 6112)	Novel Protein sim. GBank gj728837 sp P39194 ALU7_HUMAN - !!! SUBFAMILY SQ WARNING ENTRY !!!		glycoprotein	264488, 264569, 18108394, 52646842, 22278997, 22278998, 22278999, 264259, 66714117, 29331826, 29331827, 35696052, 264508, 264509, 264905, 264906, 264807, 264908, 264909, 265006, 264512, 265007, 265008, 265009, 264910, 33657402, 55812038, 264596, 264758, 265010, 265011, 265017, 265019, 264760, 18108351, 264762, 264763, 264764, 264288, 264766, 264687, 18108357, 264768, 264769, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 264691, 264693, 33657109, 18108370, 264628, 264629, 18108374, 55811576, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 18108381, 83373044, 18108385, 22279000, 22279002, 264563, 264564, 264565, 264566, 264486, 264567



3057	95412746 (6113, 6114)	Novel Protein sim. GBank gi 3878119 emb CAA88860  - (Z48068) similar to GTP-binding protein; cDNA EST EMBL:M89111 comes from this gene; cDNA EST EMBL:D27709 comes from this gene; cDNA EST EMBL:D73788 comes from this gene; cDNA EST yk353...		struct	264508, 264905, 264907, 264908, 264909, 264510, 264512, 264910, 264592, 264594, 264767, 18108374, 264635, 264555, 264637, 264639, 264563, 264564, 264565, 264486
3058	79846226 (6115, 6116)	Novel Protein sim. GBank		UNCLASSIFIED	264693
3059	87629425 (6117, 6118)	gi 4588034 gb AAD25962.1 AF09287 - (AF092878) zinc RING finger protein SAG [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	22278995, 22278996, 22278997, 22278998, 22278999, 264490, 264259, 29331824, 29331825, 29331827, 35696052, 29331828, 265007, 60433438, 265017, 265018, 265019, 264681, 264448, 264288, 264768, 21908785, 21908766, 21906787, 21906769, 29148629, 29148784, 265022, 52644150, 18108370, 264636, 18108385, 264563, 264567
3060	79346891 (6118, 6120)			UNCLASSIFIED	264567
3061	87740964 (6121, 6122)			UNCLASSIFIED	264112, 52644296, 21906768, 33657023, 263974, 18108385
3062	87619465 (6123, 6124)	Novel Protein sim. GBank gi 4454690 gb AAD20963  - (AF070657) glutathione S-transferase subunit 13 homolog [Homo sapiens]		transferase	264908, 265008, 18108351, 264566
3063	80078023 (6125, 6126)	Novel Protein sim. GBank gi 2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	18108359, 264558
3064	91241526 (6127, 6128)	Novel Protein sim. GBank gi 4240315 dbj BAA74936.1  - (AB020720) KIAA0913 protein [Homo sapiens]	Contains protein domain (PF00403) - Heavy-metal-associated domain	UNCLASSIFIED	52646355, 52646842, 65274572, 56182575, 56181688, 22278995, 22278996, 22278997, 22278999, 264259, 60432049, 29331824, 66714117, 264508, 264907, 264908, 56182435, 265009, 60432229, 60433438, 55812038, 52644296, 265018, 264682, 264288, 264686, 264768, 264687, 52644229, 264689, 21908768, 264691, 264692, 264693, 18108370, 18108377, 55811576, 264636, 56182323, 264558, 264639, 18108385, 22279000, 22279002
3065	91639201 (6129, 6130)	Novel Protein sim. GBank gi 5656743 gb AAD45960.1 AC00506 - (AC005067) Supported by Human EST H08032.1 (NID:g872854), mouse EST AA870042.1 (NID:g2965487), and gencan [Homo sapiens]		UNCLASSIFIED	22278996, 22278998, 264093, 264094, 264095, 29331824, 60424269, 66714117, 264100, 264907, 265007, 264591, 60432229, 264593, 265011, 265019, 18108351, 264766, 264767, 21908765, 21908768, 264693, 20281069, 22279000, 22279002, 264482, 264566, 264567

3066	91224437 (6131, 6132)	Novel Protein sim. GBank gi 4884268 embjCAB43245.1  - (AL050028) hypothetical protein [Homo sapiens]		UNCLASSIFIED	18108397, 22278995, 56994075, 22278996, 264905, 66712502, 265006, 264512, 264910, 264758, 60174639, 264760, 18108351, 264764, 264683, 18108359, 264692, 18108364, 18108368, 18108370, 18108377, 18108379, 60170394, 264567
3067	95422551 (6133, 6134)	Novel Protein sim. GBank gi 4689258 gb AAD27832.1 AF12185 - (AF12185) sorting nexin 9 [Homo sapiens]	Contains protein domain (PF00787) - sirud PX domain		264488, 264489, 35696286, 22278996, 56994075, 264259, 29331822, 29331825, 35696052, 29331828, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264112, 264510, 264511, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264757, 264595, 264596, 264758, 265010, 265011, 87168559, 264601, 264602, 264603, 264604, 264605, 265019, 264760, 264762, 264448, 264763, 264764, 264288, 264369, 264766, 264768, 264687, 264769, 264689, 21906765, 21906767, 21906768, 35695917, 265020, 265021, 264534, 52644150, 264691, 33657023, 264693, 264628, 60431528, 263977, 35695855, 264630, 264631, 264634, 264835, 264836, 264637, 264638, 264639, 83373044, 56526486, 87168518, 22279000, 22279002, 264563, 264483, 264564, 264565, 264566, 264567, 264486
3068	85360651 (6135, 6136)	Novel Protein sim. GBank gi 3878119 embjCAA88860  - (Z49068) similar to GTP-binding protein; cDNA EST EMBL:M89111 comes from this gene; cDNA EST EMBL:D27709 comes from this gene; cDNA EST EMBL:D27708 comes from this gene; cDNA EST EMBL:D73788 comes from this gene; cDNA EST yk353...	Contains protein domain (PF01926) - sirud GTPase of unknown function		22278996, 56994075, 22278998, 22278999, 264259, 264107, 264905, 29331830, 52644045, 264110, 60170831, 264592, 264594, 33657402, 21906754, 33109954, 87168474, 87168559, 265017, 264448, 264764, 264683, 264766, 52644229, 21906765, 21906766, 21906768, 21906769, 60170815, 33657023, 18108370, 18108376, 264634, 264557, 60170394, 56182323, 18108385, 87168518, 22279000, 264482

3070	94318173 (6139, 6140)	Novel Protein sim. GBank gi 3877788 emb CAB05527  - (Z83110) cDNA EST yk472b5.3 comes from this gene; cDNA EST yk474a7.3 comes from this gene; cDNA EST yk472b5.5 comes from this gene; cDNA EST yk468c10.3 comes from this gene; cDNA EST yk468c10.5 comes from this gene; cDNA EST EM...		synthase	264488, 22278994, 22278995, 22278996, 56994075, 22278997, 22278999, 264259, 29331822, 29147620, 29331824, 66714117, 29331826, 29146498, 29146499, 66712502, 29331830, 52644045, 56182435, 264511, 265007, 264512, 264910, 60170831, 264592, 264758, 33109954, 21906754, 87168474, 265019, 18108351, 264448, 264683, 264288, 52644229, 264689, 21906765, 21906768, 21906767, 21906769, 35695917, 265020, 265021, 60170815, 52644150, 264691, 33657023, 27486261, 27486264, 264628, 18108370, 18108377, 55811576, 35695855, 264634, 264635, 18108381, 60170394, 55182323, 264558, 83373044, 18108385, 18108387, 56528486, 264404, 264563, 264566
3071	94325573 (6141, 6142)	Novel Protein sim. GBank gi 4502425 ref NP_001709.1 pBMP6 - bone morphogenetic protein 6 precursor	Contains protein domain (PF00085) - Igf Thioresoxin		264488, 65274572, 18108398, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 66714117, 29331826, 35696052, 29331828, 29146498, 29146489, 264907, 264908, 29331830, 264909, 52644045, 56182435, 265006, 265007, 264512, 265008, 265009, 60170831, 60432228, 264592, 60433356, 33657402, 60433438, 33109954, 52644296, 87168474, 265010, 265017, 264681, 264288, 264685, 264766, 264687, 264769, 264689, 21906765, 21906768, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 60170815, 52644150, 264690, 264691, 264692, 33657023, 264693, 33657109, 263971, 18108374, 18108377, 35696423, 55811576, 65274791, 35695855, 264630, 264635, 264636, 264557, 60170394, 83373044, 60432113, 22279000, 22279002, 264563, 264564, 264565, 264566, 264567, 55182575, 29331822, 29331824, 29331825, 29146498, 264908, 52644045, 56182435, 265009, 60433438, 55812038, 18108351, 264683, 264369, 52644229, 52644150, 33657023, 264693, 33657109, 18108374, 55811576, 65274791, 264555, 56182323, 60432113, 264564
3072	95115892 (6143, 6144)	Novel Protein sim. GBank gi 1263289 (U47856) - fibroin-4 [Araneus diadematus]		transcriptfactor	

3073	86147248 (6145, 6146)	Novel Protein sim. GBank gi134840 sp P23282 CORB_HUMAN - CORNIFIN B (SMALL PROLINE-RICH PROTEIN IB) (SPR-IB) (14.9 KD PANCORNULIN)		UNCLASSIFIED	264769
3074	88099351 (6147, 6148)	Novel Protein sim. GBank gi13419847 (AC004982) - similar to yeast hypothetical protein ybk4; similar to P38164 (PID:g586461) [Homo sapiens]		UNCLASSIFIED	264488, 265019, 264448, 264288, 21906767, 264693, 18108368, 18108370, 18108374, 264567
3075	88095762 (6149, 6150)	Novel Protein sim. GBank gi4557349 ref NP_000456.1 pBARD - BRCA1 associated RING domain 1	Contains protein domain (PF00023) - Ank repeat	homeobox	264509, 264907, 264689, 264693, 56526486
3076	87819219 (6151, 6152)			UNCLASSIFIED	18108398, 29331822, 29331827, 60432228, 265017, 264691, 264693
3077	88734277 (6153, 6154)	Novel Protein sim. GBank gi3023956 sp Q00808 HET1_PODAN - VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	65274572, 35698052, 284511, 60170831, 87168474, 264369, 35695917, 33657182, 27486284, 33657349, 35695763, 35695855, 264639
3078	88095355 (6155, 6156)	Novel Protein sim. GBank gi3900850 (AC004994) - similar to KIAA0600; similar to d1026456 (PID:g3043724) [Homo sapiens]			22278002
3079	87821893 (6157, 6158)	Novel Protein sim. GBank gi3875410 emb CAB02876  - (Z81052) Similarity to Yeast ABC1P protein (SW/ABC1_YEAST); cDNA EST yk229g8.3 comes from this gene; cDNA EST yk229g8.5 comes from this gene [Caenorhabditis elegans]		transport	29331824, 29331826, 264758, 55811386, 265017, 55811150, 52644228, 21906768, 265020, 265021, 264693, 18108376, 264631, 52644332, 22279002
3080	95298274 (6159, 6160)	Novel Protein sim. GBank gi5257221 gb AAD41265.1  - (AF117887) protein arginine methyltransferase [Mus musculus]		interferon	264488, 52644507, 22278996, 22278998, 264490, 264259, 29331824, 66714117, 29331825, 29331826, 29331827, 29331828, 29146499, 264508, 264905, 264828, 52644045, 56182435, 265006, 264591, 264596, 21906754, 60174639, 265010, 264882, 264448, 264763, 264683, 264764, 264288, 264685, 264769, 264688, 264689, 21906765, 21906767, 21906769, 55811957, 35695917, 265020, 60170615, 52644150, 264692, 33657023, 264693, 65274620, 33657109, 27486261, 35695763, 264628, 18108370, 65274791, 264558, 56182323, 60170394, 264482, 264565, 264484
3081	88094864 (6161, 6162)	Novel Protein sim. GBank gi728831 sp P39188 ALU1_HUMAN - !!!!! ALU SUBFAMILY J WARNING ENTRY !!!!!		UNCLASSIFIED	18108398, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 264595, 264758, 265011, 265018, 264760, 264761, 264763, 264764, 18108354, 264685, 264766, 264628, 264630, 264631, 264632, 264634, 264635, 264555, 264638, 18108382, 18108385, 264563, 264565, 264566
3082	80310121 (6163, 6164)				264764, 55811957, 264555, 264564

3083	88095756 (6165, 6166)	Novel Protein sim. GBank gi 868241 (U28488) - C56C10.3 gene product [Caenorhabditis elegans]		UNCLASSIFIED	284488, 284259, 29331824, 284106, 265008, 284591, 264592, 21906754, 264288, 284767, 21906768, 21906769, 29148784, 284691, 264632, 22279000
3084	87448568 (6167, 6168)	Novel Protein sim. GBank gi 476774 pir A37475 - probable structural component p38 - borna disease virus			22278995, 60432289, 35696052, 264805, 264906, 264907, 264908, 264909, 265006, 265007, 264910, 264593, 264595, 264758, 264369, 264288, 264766, 35695917, 265020, 18108374, 35696423, 264631, 264556, 264555, 264586, 264587, 264486
3085	87795781 (6169, 6170)	Novel Protein sim. GBank gi 2565037 (U80741) - CAGH44 [Homo sapiens]		UNCLASSIFIED	265011, 284681
3086	87769942 (6171, 6172)	Novel Protein sim. GBank gi 3894189 (AC005662) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	22278998, 264092, 264259, 29331822, 29331825, 264108, 264112, 18108351, 264687, 263967, 263974, 55810764, 263981, 18108385, 284487
3087	87462888 (6173, 6174)				52646365, 56994075, 22278997, 22278998, 29331824, 29331825, 35696052, 60433438, 33109954, 21906754, 52646317, 285017, 264682, 264369, 264684, 21906767, 21906768, 265020, 264691, 33657023, 33657109, 52645129, 33657182, 27486262, 35695855, 87168518
3088	91224441 (6175, 6176)	Novel Protein sim. GBank gi 3355304 (AF001549) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264591
3089	95361242 (6177, 6178)	Novel Protein sim. GBank gi 4689146 gb AAD27782.1 AF07704 - (AF077049) lambda-crystallin [Homo sapiens]	Contains protein domain (PF00725) - dehydrogenase 3-hydroxyacyl-CoA dehydrogenase	dehydrogenase	18108397, 65274572, 56182575, 56181686, 56994075, 35696286, 22278997, 22278998, 264259, 29331824, 29331825, 29331826, 29331828, 264907, 29331830, 264909, 56182435, 264510, 265007, 60170831, 60432229, 21906754, 55811386, 265017, 265018, 265019, 264760, 55811150, 264288, 264766, 56181562, 21906765, 21906766, 21906767, 21906768, 265021, 60170615, 27486262, 18108370, 60431528, 35696423, 264556, 264559, 60432113, 264486

3090	95342371 (6179, 6180)	Novel Protein sim. GBank gij1354050 (U47024) - MEM3 [Mus musculus]	UNCLASSIFIED	60424179, 52645156, 65274572, 56182575, 56181686, 22278995, 35696286, 56994075, 22278996, 22278998, 22278999, 264259, 29331822, 56182181, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 33656970, 264906, 264908, 52644045, 264828, 265006, 265007, 265008, 60170831, 60432229, 60433356, 33657402, 55812038, 264758, 21906754, 33109954, 52646317, 55811386, 52844296, 87168474, 265011, 87168559, 265017, 265018, 265019, 55811150, 18108351, 264681, 264448, 264288, 264369, 18108357, 264768, 52644229, 56181562, 21906764, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265022, 60170615, 264690, 52644150, 264691, 33657023, 18108365, 65274620, 33657109, 18108368, 33657182, 27486261, 27486265, 35695763, 18108374, 18108376, 55810764, 35696423, 55811576, 65274791, 35695855, 264557, 56182323, 83373044, 18108387, 18108388, 87168518, 22278000, 22279002, 264563, 264482
3091	95317424 (6181, 6182)	Novel Protein sim. GBank gij3873932[emb]CAB01859] - (Z79596) Similarity to Bovine asparyl beta hydroxylase (TR:G162694); cDNA EST EMBL:D27916 comes from this gene; cDNA EST EMBL:D27915 comes from this gene; cDNA EST EMBL:D64881 comes from this gene; cDNA EST EMBL:D68139 comes f...	UNCLASSIFIED	35695286, 29331822, 35696052, 264508, 264509, 264905, 264906, 264908, 264909, 264510, 264758, 265010, 265011, 264683, 264685, 264766, 264768, 264769, 264693, 264628, 35696423, 35695855, 264632, 264635, 264639, 264482, 264563, 264486

3092	95314592 (6183, 6184)	Novel Protein sim. GBank gi1710756 sp P15880 RS2_HUMAN - 40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN)	Contains protein domain (PF00333) - Ribosomal protein S5	264488, 60424179, 18108396, 22278995, 56994075, 22278996, 35696286, 22278997, 22278998, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 29331828, 29146498, 29146499, 284508, 284509, 264905, 264906, 264907, 29331830, 264908, 284909, 264113, 264510, 264511, 265006, 264512, 265007, 265008, 264910, 265009, 60170831, 264591, 264592, 60431735, 264593, 264594, 60433438, 264595, 264758, 21906754, 265010, 265011, 264601, 264602, 265017, 264603, 264604, 265018, 264805, 265019, 264760, 264762, 264681, 18108351, 264763, 264682, 264448, 264764, 264683, 264288, 264369, 264765, 264766, 264686, 264767, 264687, 264768, 264769, 264688, 21906764, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 29148629, 29148784, 35695917, 265020, 265021, 264534, 60170615, 264690, 264691, 264692, 65274820, 33657109, 27486262, 264628, 264629, 18108374, 263978, 18108377, 35696423, 264630, 264631, 264632, 264634, 264635, 264555, 264636, 264637, 264556, 264638, 264557, 264558, 264639, 60170394, 18108385, 264259, 29331824, 35696052, 264905, 265006, 60432229, 60431735, 264684, 264369, 264288, 264766, 21906767, 35696423, 83373044, 18108385
3093	94318457 (6185, 6186)	Novel Protein sim. GBank gi15002587 emb CAB44347.1  - (Y17454) LSF1 protein [Homo sapiens]	UNCLASSIFIED	18108398, 264259, 60432289, 29331827, 264511, 264763, 264288, 264767, 265022, 264691, 264693, 65274791, 56182323, 264564, 264565
3094	94316675 (6187, 6188)	Novel Protein sim. GBank gi1400734 sp P31044 PBP_RAT - PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN (23 KD MORPHINE-BINDING PROTEIN) (P23K)	Contains protein domain (PF01161) - Phosphatidylethanolamine-binding protein	

3095	84848182 (6189, 6180)	Novel Protein sim. GBank gi 4877759 gb AAD31421.1 AF12444 - (AF124440) MAGE tumor antigen D1 [Homo sapiens]	Contains protein domain (PF01454) - MAGE family	UNCLASSIFIED	18108397, 56182575, 22278995, 35696286, 56994075, 22278997, 22278999, 264259, 60432049, 66714117, 29331825, 60432289, 35696052, 33656970, 29146499, 264508, 264905, 264509, 29331830, 264909, 264510, 264511, 264512, 265007, 265008, 265009, 60170831, 264758, 21906754, 85658542, 265010, 265011, 87168559, 265017, 265018, 265019, 264760, 264681, 264682, 264683, 264764, 264369, 264288, 264886, 264788, 264769, 264689, 21906765, 21906766, 21906767, 55811957, 35695917, 265020, 265021, 265022, 52644150, 264691, 264692, 33657023, 264693, 263972, 18108376, 55811576, 35696423, 264952, 60170394, 264639, 83373044, 18108385, 18108387, 65274727, 87168518, 60432113, 264482, 264563, 264564, 264566, 264487, 18108391, 22278995, 22278996, 22278997, 22278999, 29331824, 29331825, 29331826, 29331827, 33656970, 264905, 264908, 265008, 264910, 33657402, 265011, 265017, 265018, 264369, 21906766, 21906767, 21906768, 35695917, 265020, 60170615, 264691, 264692, 264693, 27486261, 27486262, 18108370, 60431528, 264634, 264636, 264639, 22279000, 264566
3096	87756128 (6191, 6192)	Novel Protein sim. GBank gi 3882221 dbj BAA34470.1  - (AB018293) KIAA0750 protein [Homo sapiens]	Contains protein domain (PF00307) - Calponin homology (CH) domain	struct	264488, 29331822, 29331825, 60432289, 29331826, 35696052, 29331828, 29331830, 264594, 55812038, 33109854, 33657084, 87168474, 87168559, 52644229, 21906765, 21906767, 18108376, 35696423, 52644332, 264638, 60432113, 22279002
3097	88264895 (6183, 6194)	Novel Protein sim. GBank gi 4468288 emb CAB37981  - (AL022395) dJ273N12.1 (PUTATIVE protein based on EST matches) [Homo sapiens]	Contains protein domain (PF00646) - F-box domain.	UNCLASSIFIED	264488, 29331822, 29331825, 60432289, 29331826, 35696052, 29331828, 29331830, 264594, 55812038, 33109854, 33657084, 87168474, 87168559, 52644229, 21906765, 21906767, 18108376, 35696423, 52644332, 264638, 60432113, 22279002
3098	80258024 (6195, 6196)				264634, 264637, 264565
3099	91243325 (6197, 6198)	Novel Protein sim. GBank gi 303603 dbj BAA02145.1  - (D12621) cytochrome P-450 TBV [Homo sapiens]		cyto450	264488, 35696286, 29331822, 29331824, 29331825, 29331827, 265007, 265008, 265010, 265011, 265018, 265019, 18108357, 21906766, 265020, 265022, 55811576,
3100	87602421 (6199, 6200)	Novel Protein sim. GBank gi 1083764 pir B48013 - proline- rich proteoglycan 2 precursor, parolid - rat		UNCLASSIFIED	58182323, 22279002, 264563
3101	79602134 (6201, 6202)			UNCLASSIFIED	29331825, 60432289, 35696052, 264910, 60432229, 264592, 264288, 264693, 263967, 264635
				UNCLASSIFIED	264908, 264693, 264628, 264630, 264632



3102	91220892 (6203, 6204)	Novel Protein sim. GBank gi 5305708 gb AAD41781.1 AF12853 - (AF128536) cytoplasmic phosphoprotein PACSIN2 [Homo sapiens]	Contains protein domain (PF00018) - struct SH3 domain	35696286, 22278996, 22278999, 29331827, 35696052, 264908, 264511, 285008, 60170831, 60433356, 33109954, 18108351, 264684, 264689, 21906767, 60170615, 264692, 33657023, 264638, 22279000, 264482, 264564 35695917, 264565
3103	90838004 (6205, 6206)	Novel Protein sim. GBank gi 464564 sp P35292 RB17_MOUSE - RAS-RELATED PROTEIN RAB-17	UNCLASSIFIED	
3104	87340833 (6207, 6208)	Novel Protein sim. GBank gi 5032207 ref NP_005696.1 pTSSC - tumor-suppressing STF cDNA 6	UNCLASSIFIED	264259, 264684, 264532, 33657182, 264558
3105	94148603 (6209, 6210)			22278997, 264259, 29331824, 35696052, 29331828, 264508, 264509, 264905, 264906, 264907, 264908, 264511, 264910, 264591, 264594, 264758, 264760, 264681, 264762, 264764, 264288, 264766, 264768, 264687, 264769, 21906766, 21906768, 35695917, 33657023, 264692, 264693, 264628, 264629, 35695855, 264630, 264631, 264632, 264634, 264635, 264637, 264638, 264639, 83373044, 264404, 22279002, 264563, 264585, 264566, 264486, 264587
3106	95361416 (6211, 6212)	Novel Protein sim. GBank gi 1938574 (U97190) - B0025.2 gene product [Caenorhabditis elegans]		22278996, 22278997, 22278998, 22278999, 264092, 264093, 264094, 29331822, 264906, 264907, 264908, 52644045, 56182435, 264112, 265008, 265009, 55812038, 265017, 265018, 264683, 264686, 264687, 264788, 52644228, 21906765, 21906768, 21906769, 55811957, 265020, 265022, 264690, 52644150, 264692, 264693, 18108370, 18108377, 55811576, 56182323, 18108385, 18108388, 22279000, 264583
3107	95343272 (6213, 6214)	Novel Protein sim. GBank gi 3341441 emb CAA76851  - (Y17794) winged-helix transcription factor [Gallus gallus]		22278995, 22278996, 35696286, 22278997, 22278999, 264091, 264093, 264259, 29331822, 29331825, 29331826, 60432289, 29331827, 29331828, 33656970, 264105, 264512, 265009, 60433356, 60433438, 265011, 265017, 265018, 21906785, 21906766, 21906767, 21906769, 265021, 264691, 33657109, 27486261, 27486265, 18108370, 263972, 18108374, 55811576, 18108385, 56526488, 264482, 264487, 56182435, 264288, 264690, 264584
3108	87340635 (6215, 6216)	Novel Protein sim. GBank gi 5032207 ref NP_005696.1 pTSSC - tumor-suppressing STF cDNA 6	UNCLASSIFIED	

3109	94318461 (6217, 6218)	Novel Protein sim. GBank gii5002587[emb][CAB44347.1] - (Y17454) LSFRT protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	struct	264490, 264908, 265007, 264910, 264593, 264683, 264684, 264687, 21906767, 21906768, 264693, 18108370, 264629, 18108374, 264632, 264638, 22279000
3110	95090716 (6219, 6220)	Novel Protein sim. GBank gii1076211[pir][S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii]		UNCLASSIFIED	264488, 65274572, 22278995, 22278997, 60432049, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 264908, 264510, 265006, 265007, 265008, 265009, 60432229, 33657402, 60433356, 265011, 87168559, 264600, 265017, 265018, 265019, 18108351, 264288, 264369, 21906768, 21906767, 21906768, 265020, 60170615, 264693, 65274620, 18108370, 264638, 18108384, 22278000, 264563, 18108390
3111	87754512 (6221, 6222)	Novel Protein sim. GBank gii3282231 (U75454) - C2H2 type zinc finger protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264488, 18108398, 66712502, 265017, 265018, 265019, 264448, 21906767, 265020, 33657023, 18108365, 18108368, 35698423, 52644332, 18108385, 18108388
3112	88043639 (6223, 6224)	Novel Protein sim. GBank gii3900848 (AC005023) - match to EST AA361117 (NID:q2013436) [Homo sapiens]	Contains protein domain (PF00046) - Homeobox domain	homeobox	
3113	88207098 (6225, 6226)	Novel Protein sim. GBank gii2459910 (AF005856) - anon2A5 [Drosophila yakuba]		im7	18108397, 22278999, 264259, 29331824, 35698052, 264907, 264757, 60433438, 87168559, 264763, 264448, 18108354, 264288, 21906767, 21906769, 35695917, 264690, 264691, 264692, 264693, 18108365, 18108381, 18108384, 18108385, 18108386, 87168518, 22278000, 22278002
3114	79843167 (6227, 6228)	Novel Protein sim. GBank gii4966270[gb][AAB52261.2] - (U97002) similar to acyl-CoA dehydrogenases and epoxide hydrolases; Pfam domain PF00441 (Acy-CoA_dh), Score=57.4, E-value=1.7e-16, N=2; contains similarity to Pfam domain PF00702 (Hydrolase), Score=57.4, E-value=1e-13, N=1 [C...]	Contains protein domain (PF00702) - haloacid dehalogenase-like hydrolase	hydrolase	264909, 56182435, 264910, 21906754
3115	94117896 (6229, 6230)	Novel Protein sim. GBank gii5032225[re][NP_005676.1]pWBSC - Williams-Beuren syndrome chromosome region 11		transcriptfactor	60424179, 56182575, 264259, 29331824, 60424269, 29331826, 66712502, 264510, 265007, 60431735, 60433356, 55812038, 55811386, 265018, 264288, 264689, 21906769, 264691, 33657023, 264693, 60431528, 263974, 60431850, 56182323, 264559, 22279000, 22279002
3116	79842855 (6231, 6232)			UNCLASSIFIED	264905, 264758, 21906784, 264680
3117	8771288 (6233, 6234)			UNCLASSIFIED	264510, 265011, 18108351, 264288, 264689, 264691, 18108368, 18108372, 263981, 264558, 264564

3118	94665848 (6235, 6236)	Novel Protein sim. GBank gij3880563jemb[CAB01444.1] - (Z78018) predicted using GeneFinder; similar to serine/threonine kinase; cDNA EST yk353d10.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00008) - EGF-like domain	Igf	52645156, 52646842, 65274572, 56182575, 22278995, 56994075, 22278996, 35698286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 66714117, 29331826, 29331827, 35696052, 29331828, 264905, 264908, 29331830, 52644045, 56182435, 264510, 264511, 265007, 265008, 265009, 264757, 52646317, 21906754, 33657084, 52644286, 87168474, 87168559, 265017, 265018, 264605, 265019, 264762, 264448, 264682, 264684, 264288, 264766, 56181562, 21906765, 21906766, 21906768, 21906769, 265020, 265022, 264690, 52644150, 264691, 33657023, 264693, 33657109, 33657349, 264628, 18108370, 60431528, 18108374, 35698423, 65274791, 60170394, 83373044, 87168518, 22279000, 22279002, 264486, 265006, 264288
3119	85728786 (6237, 6238)		Contains protein domain (PF00328) - Histidine acid phosphatase		
3120	87344040 (6239, 6240)	Novel Protein sim. GBank gij5019819gib AAD37863.1 AF14315 - (AF143152) putative NADH oxidoreductase complex I subunit [Caenorhabditis elegans]		UNCLASSIFIED	264488, 264509, 264510, 264511, 264512, 264288, 264486
3121	94110735 (6241, 6242)	Novel Protein sim. GBank gij4501877 red NP_001088.1 pACR  - acrosin			52644507, 52645156, 52646385, 52646842, 22278994, 56994075, 22278996, 22278999, 264259, 29331824, 29331827, 35696052, 52644045, 265008, 52646317, 87168474, 87168559, 21906765, 52644150, 33657023, 18108374, 264637
3122	11814528 (6243, 6244)			UNCLASSIFIED	264638
3123	88083003 (6245, 6246)	Novel Protein sim. GBank gij2439517 (AC002563) - putative RHO/RAC effector protein; 95% similarity to P49205 (PID:g1345860) [Homo sapiens]	Contains protein domain (PF00780) - CNH domain	kinase	18108392, 29331822, 29331824, 29331825, 264905, 265007, 55812038, 265019, 18108351, 264682, 264288, 264766, 21906764, 21906765, 21906768, 21906769, 55811957, 18108365, 18108366, 27486265, 18108374, 18108381, 18108384, 22279000, 22279002, 264482
3124	87786899 (6247, 6248)			UNCLASSIFIED	264905
3125	91216607 (6249, 6250)	Novel Protein sim. GBank gij4980826 gib AAD35412.1 AE001714) oxidoreductase, short chain dehydrogenase/reductase family [Thermotoga maritima]	Contains protein domain (PF00106) - short chain dehydrogenase	dehydrogenase	56181686, 264259, 66714117, 60433289, 29331826, 29331827, 264907, 264908, 264828, 265009, 60433356, 33657402, 60433438, 264758, 18108351, 264288, 29148827, 29148629, 33657023, 33657109, 18108382, 56526486

3128	95337205 (6251, 6252)			UNCLASSIFIED	22278999, 264490, 264259, 60432049, 29331822, 60432289, 29146498, 52844045, 56182435, 265009, 60433438, 265010, 87168559, 265017, 265018, 55811150, 264763, 264683, 264369, 264685, 29148629, 33657023, 264693, 33657109, 18108374, 55811576, 18108385, 60432113, 22279002, 35566286, 22278996, 22278999, 29331826, 264908, 60433438, 87168559, 264604, 21906765, 21906769, 33657023, 33657349, 264629, 18108374, 18108377, 22279000, 22279002
3127	91639233 (6253, 6254)	Novel Protein sim. GBank gj 2828280 emb CAA16694.1  - (AL021687) putative protein [Arabidopsis thaliana]			
3128	87674330 (6255, 6256)	Novel Protein sim. GBank gj 3885828 (AF090133) - lin-7-A [Rattus norvegicus]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).	misc_channel	22278996, 264259, 52844045, 265008, 21906754, 265017, 265018, 21906768, 18108376, 18108387, 22278999, 22279002
3129	87755412 (6257, 6258)	Novel Protein sim. GBank gj 3135273 (AC003058) - hypothetical protein [Arabidopsis thaliana]	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	56182575, 264259, 29331825, 29331828, 52844045, 56182435, 60433356, 264600, 264682, 264763, 264764, 264369, 264288, 264686, 55811957, 264692, 33657023, 33657109, 60432113, 264564, 264566, 264636
3130	14983960 (6259, 6260)	Novel Protein sim. GBank gj 3329465 (AF064553) - NSD1 protein [Mus musculus]			
3131	95351469 (6261, 6262)	Novel Protein sim. GBank gj 1848277 (U86136) - telomerase-associated protein TP-1 [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	56182575, 264259, 29331824, 264907, 56182435, 264594, 60433438, 55812038, 33109954, 21906754, 33657084, 87168474, 264448, 264766, 21906769, 55811957, 265020, 265021, 265022, 60170615, 33657023, 33657109, 33657182, 27486261, 33657349, 65274791, 60170394, 56182323, 83373044, 87168518, 264564

3132	95415459 (6263, 6264)	Novel Protein sim. GBank gi 4680647 gb AAD27713.1 AF13293 - (AF13293) CGI-03 protein [Homo sapiens]	Contains protein domain (PF00789) - UBX domain	- ubiquitin	52644507, 52646842, 52646365, 55274572, 56182575, 22278994, 22278995, 35696286, 56994075, 22278996, 22278997, 22278998, 22278999, 60432049, 52645080, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264508, 52644045, 56182435, 264910, 60170831, 60432228, 60433356, 33657402, 55812038, 52646317, 21906754, 52644296, 85658542, 87168559, 265017, 265018, 265019, 264448, 264288, 264369, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 60170815, 52644150, 264692, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486265, 33657349, 35695763, 18108374, 18108376, 55811576, 35695855, 18108385, 18108387, 56526486, 87168518, 60432113, 22279002
3133	87379414 (6265, 6266)	Novel Protein sim. GBank gi 4507613 ref NP_003738.1 pTNKS - TANKYRASE		polymerase	22278994, 22278998, 264905, 265006, 265007, 87168559, 264760, 21906767, 18108374, 22278000, 22279002, 264563
3134	94848816 (6267, 6268)	Novel Protein sim. GBank gi 1729827 sp P54833 TALA_DICDI - FILOPODIN (TALIN HOMOLOG)			264595, 264369, 264685, 264628, 264566
3135	86389356 (6269, 6270)	Novel Protein sim. GBank gi 3093478 (AF012927) - fibrogen-binding protein [Streptococcus equi]		struct	22278998, 264095, 28331826, 33657402, 18108348, 263974
3136	84845838 (6271, 6272)	Novel Protein sim. GBank gi 627101 pir S44092 - probable carrier protein c2 - Caenorhabditis elegans	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	22278998, 264259, 264828, 265008, 265008, 60433438, 265019, 264764, 264288, 264769, 264689, 265020, 27486262, 263972, 65274791, 264557, 264558
3137	88257947 (6273, 6274)	Novel Protein sim. GBank gi 3342730 (AC005331) - R31341_1 [Homo sapiens]		UNCLASSIFIED	22278995, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331825, 29331826, 29331827, 29331828, 264510, 265008, 21906754, 87168474, 265011, 87168559, 265017, 265018, 265019, 18108351, 264682, 264769, 21906765, 21906766, 21906767, 21906769, 55811957, 35695917, 265020, 265021, 52644150, 18108370, 18108374, 22279000, 22279002, 264482, 264486

3138	94130186 (6275, 6276)	Novel Protein sim. GBank gll4406759lgbjAAD20070] - (AC006838) hypothetical protein [Arabidopsis thaliana]			264569, 264488, 264907, 264511, 264593, 33109954, 87168559, 264681, 264684, 264686, 264688, 264687, 264768, 264688, 264689, 264691, 264692, 264693, 33657109, 264631, 264634, 264635, 264636, 264637, 60170394, 83373044, 18108385, 18108388, 60432113, 22278000, 22278002
3139	87325503 (6277, 6278)	Novel Protein sim. GBank gll228938prll1814452C - Hyp-rich glycoprotein [Zea diploperennis]	UNCLASSIFIED		265018, 265019, 21906765, 265020, 264636, 264557
3140	91222692 (6278, 6280)	Novel Protein sim. GBank gll932lombjCAA37773] - (X53744) 68kDa subunit of signal recognition particle [Canis familiaris]	struct		22278995, 56994075, 35696286, 264908, 264909, 60433356, 21906754, 52644296, 87168474, 87168559, 264683, 264288, 264685, 264686, 265022, 264693, 27486262, 35695855, 264630, 264555, 264566
3141	87323564 (6281, 6282)	Novel Protein sim. GBank gll3213227 (AF035209) - putative v-SNARE VII1a [Mus musculus]	UNCLASSIFIED		56182575, 35696286, 28331828, 264909, 265009, 265018, 18108351, 264369, 21906766, 29148627, 265020, 264628, 264629, 264631, 18108385
3142	95419028 (6283, 6284)	Novel Protein sim. GBank gll2498197lspIQ95245JC561_PIG - CYTOCHROME B561 (CYTOCHROME B-561)	cytochrome		52645156, 52646365, 22278995, 35696286, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331827, 29146499, 56182435, 265007, 60170831, 60432229, 33657402, 264595, 60433438, 264758, 21906754, 264288, 264766, 264687, 52644229, 21906765, 21906767, 21906768, 60170815, 52644150, 65274620, 33657109, 35695763, 18108370, 18108376, 65274791, 35695855, 264631, 264557, 87168518, 60432113, 22279000
3143	95351475 (6285, 6286)	Novel Protein sim. GBank gll5420387lombjCAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED		264488, 56182575, 22278996, 22278998, 22278999, 28331822, 29331824, 60432289, 35696052, 29331828, 264508, 264905, 264906, 264907, 264908, 264909, 52644045, 56182435, 264511, 264512, 265008, 264910, 60432229, 33657402, 60433356, 60433438, 55812038, 265011, 265019, 264760, 264763, 264448, 264764, 264684, 264288, 264685, 264686, 264768, 264689, 21906765, 21906766, 21906767, 21906769, 35695917, 264690, 33657023, 264693, 263967, 33657109, 264628, 264629, 18108374, 263976, 55811576, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264558, 87168518, 60432113, 22279000, 22279002, 264563, 264566, 264486

3144	95336329 (6287, 6288)	Novel Protein sim. GBank gij4884468[embjCAB43322.1] - (AL050225) hypothetical protein [Homo sapiens]			264488, 18108396, 22278996, 35696286, 22278997, 22278999, 29331826, 29331827, 35696052, 29331828, 264106, 265006, 265007, 265009, 33657402, 85658542, 265011, 18108351, 264448, 264369, 21906765, 21906766, 21906767, 265020, 265021, 52844150, 27486261, 18108370, 18108374, 35696423, 56182323, 83373044, 22279000, 22279002, 264587
3145	86611657 (6289, 6290)	Novel Protein sim. GBank gij3879709[embjCAB03330] - (Z81118) Similarity to Human endosomal protein P162 (TR:Q15075); cDNA EST EMBL:Z14487 comes from this gene; cDNA EST EMBL:Z14556 comes from this gene; cDNA EST EMBL:D27011 comes from this gene; cDNA EST EMBL:D27015 comes from l...	UNCLASSIFIED		18108397, 29331824, 29146499, 20281100, 265006, 55812038, 265010, 21906766, 29148627, 21906769, 29148784, 264692, 33657023, 33657109, 35695763, 263981, 56182323, 87168518
3146	87756314 (6291, 6292)	Novel Protein sim. GBank gij2135746[pirjS69890] - mitogen inducible gene mig-2 - human	Contains protein domain (PF00169) - PH domain	struct	264259, 29331826, 29331828, 29331830, 264510, 264511, 265007, 265009, 264600, 265017, 18108351, 264448, 264369, 21906766, 265021, 264692, 33657109, 18108374, 35696423, 35695855, 60432113, 264584
3147	94848512 (6293, 6294)	Novel Protein sim. GBank gij3874270[embjCAB07315.1] - (Z92825) predicted using GeneFinder, cDNA EST yk315e12.3 comes from this gene; cDNA EST yk315e12.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00702) - haloacid dehalogenase-like hydrolase	UNCLASSIFIED	56181686, 35696286, 60432049, 264259, 56182181, 29331825, 60432289, 35696052, 56182435, 285008, 264910, 60431735, 60433356, 60433438, 265010, 264448, 264288, 265022, 33657023, 33657109, 60431528, 65274791, 264631, 56182323, 264404, 22278002
3148	95362169 (6295, 6296)	Novel Protein sim. GBank gij522532[gbjAAD40851.1]AF08310 - (AF083108) sirtuin type 3 [Homo sapiens]	UNCLASSIFIED		35696286, 35696052, 264511, 85658542, 87168474, 264784, 35696423, 264555, 264556, 264557, 264558, 83373044, 56526486, 60432113
3149	95308548 (6297, 6298)	Novel Protein sim. GBank gij4200446 (AF102777) - FYVE finger-containing phosphoinositide kinase [Mus musculus]	Contains protein domain (PF01363) - FYVE zinc finger	eph	29331822, 35696052, 264109, 29148629, 18108381
3150	87655472 (6299, 6300)	Novel Protein sim. GBank gij3378454[embjCAA76893] - (Y17850) ganglioside-induced differentiation associated protein 1 [Mus musculus]	Contains protein domain (PF00043) - Glutathione S-transferases	transferase	264259, 29331822, 29331824, 29331825, 29331827, 52646317, 264686, 35695855, 56182323, 264639
3151	87772355 (6301, 6302)	Novel Protein sim. GBank gij172591 (M63577) - SFP1 [Saccharomyces cerevisiae]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	oncogene	29331822, 265008
3152	85698108 (6303, 6304)		UNCLASSIFIED		21906754, 87168559, 264605, 21906768, 52644150, 27486264, 35696423, 22279000

3153	95317298 (6305, 6306)	Novel Protein sim. GBank gi 4895041 gb AAD32705.1 AF143957 coronin-3 [Mus musculus]	Contains protein domain (PF00400) - WD domain, G-beta repeat	struct	264488, 52646365, 35696286, 22278996, 22278997, 22278998, 60432049, 264259, 29331826, 60432289, 33656970, 264508, 264908, 33657402, 264595, 60433438, 87168474, 87168559, 264601, 265019, 264448, 264682, 264764, 264288, 264369, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 29148784, 265021, 265022, 60170615, 52644150, 264690, 264691, 33657023, 65274620, 33657109, 18108370, 35695855, 264638, 60170394, 87168518, 60432113, 22278900, 22279002, 22278998, 264259, 29331824, 66712502, 265008, 265010, 265017, 18108354, 264691, 33657023, 264693, 20281149, 18108374
3154	87718573 (6307, 6308)	Novel Protein sim. GBank gi 4680661 gb AAD27720.1 AF132945 CGI-11 protein [Homo sapiens]		ATPase-associated	29331828, 264509, 264905, 264908, 264510, 264511, 264512, 33657402, 264681, 264683, 33657023, 18108370, 264634, 264639, 18108385, 264563, 264486
3155	87762394 (6308, 6310)	Novel Protein sim. GBank gi 728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY !IIII		UNCLASSIFIED	56182575, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 29331827, 35696052, 52644045, 265007, 265009, 60170831, 60432229, 60433356, 21906754, 33109954, 87168474, 265010, 265017, 265018, 265019, 18108351, 264448, 264288, 264689, 21906766, 21906768, 21906769, 35695917, 265020, 265022, 264692, 18108370, 35696423, 56182323, 22279002
3156	87737449 (6311, 6312)	Novel Protein sim. GBank gi 5830076 gb AAD45821.1 AC00601 - (AC006017) N-acetylglucosaminyltransferase, similar to Q10473 (PID:gi1709559) [Homo sapiens]	Contains protein domain (PF00652) - Similarity to lectin domain of ricin beta-chain, 3 copies.	transferase	18108396, 264259, 29331826, 35696052, 29146498, 87168559, 265017, 264448, 264288, 264691, 18108366, 52645129, 35696423, 52644332
3157	88259577 (6313, 6314)	Novel Protein sim. GBank gi 5306064 gb AAD41895.1 AF15677 - (AF156778) ASB-3 protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	kinase	264488, 263974
3158	80034118 (6315, 6316)	Novel Protein sim. GBank gi 5306064 gb AAD41895.1 AF15677 - (AF156778) ASB-3 protein [Homo sapiens]		UNCLASSIFIED	56182575, 22278999, 29331824, 264106, 60433356, 264758, 265011, 87168559, 264448, 18108354, 264768, 21908768, 265020, 264691, 264692, 33657109, 18108374, 35696423, 264555, 60170394, 22279000
3159	84124114 (6317, 6318)	Novel Protein sim. GBank gi 5531272 emb CAB50897.1  - (AJ243800) WSC4 homologue [Kluyveromyces fragilis]		UNCLASSIFIED	18108351, 264555, 264556, 264557, 264558, 264559
3160	80221068 (6319, 6320)	Novel Protein sim. GBank gi 3930525 (AF064447) - sex-determination protein homolog Fem1a [Mus musculus]	Contains protein domain (PF00023) - Ank repeat	struct	



3161	8807411 (6321, 6322)				264486, 22278985, 22278987, 22278988, 264259, 29331822, 60432289, 28331828, 52644045, 265017, 265018, 264448, 264288, 21808764, 21808767, 285020, 18108374, 264636, 264566
------	----------------------	--	--	--	--

Table 2

Tissue ID	Tissue Name	Tissue Information	Disease Association
20281069	192xN	Protein-protein Interactions	Any
20281071	192xN	Protein-protein Interactions	Any
20281149	192xN	Protein-protein Interactions	Any
20281152	192xN	Protein-protein Interactions	Any
264111	276xN	Protein-protein Interactions	Any
264112	276xN	Protein-protein Interactions	Any
263966	384xN	Protein-protein Interactions	Any
263967	384xN	Protein-protein Interactions	Any
264110	552xN	Protein-protein Interactions	Any
18108379	SPH 52.1 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy , Congenital Adrenal Hyperplasia,
18108381	SPH 52.2 (Fetal Lung)	Fetal Lung	Cystic Fibrosis, infection, lung cancer
18108383	SPH 52.3 (B's Lymphoma- Raji)	Lymphoma derived from B cells	Blood cancers, hematopoiesis, leukemia
18108368	SPH 52.5 (Salivary Gland)	Salivary Gland	Dry mouth, infection
18108384	SPH 52.6 (Brain- Thalamus)	Thalamus	Brain cancer, head injury, obesity, neurological disorders, neuropsychiatric disorders
18108394	SPH 53.1 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy , Congenital Adrenal Hyperplasia,
18108355	SPH 53.2 (Fetal Lung)	Fetal Lung	Cystic Fibrosis, infection, lung cancer
18108359	SPH 53.3 (B's Lymphoma- Raji)	Lymphoma derived from B cells	Blood cancers, hematopoiesis, leukemia
18108361	SPH 53.4 (Mammary Gland)	Mammary Gland	Lactation disorders, breast cancer
18108362	SPH 53.5 (Salivary Gland)	Salivary Gland	Dry mouth, infection
18108366	SPH 53.6 (Brain- Thalamus)	Thalamus	Brain cancer, head injury, obesity, neurological disorders, neuropsychiatric disorders
18108354	SPH 54.1 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy , Congenital Adrenal Hyperplasia,
18108392	SPH 54.2 (Fetal Lung)	Fetal Lung	Cystic Fibrosis, infection, lung cancer
18108348	SPH 54.3 (B's Lymphoma- Raji)	Lymphoma derived from B cells	Blood cancers, hematopoiesis, leukemia
18108382	SPH 54.4 (Mammary Gland)	Mammary Gland	Lactation disorders, breast cancer
18108395	SPH 54.5 (Salivary Gland)	Salivary Gland	Dry mouth, infection
18108365	SPH 54.6 (Brain- Thalamus)	Thalamus	Brain cancer, head injury, obesity, neurological disorders, neuropsychiatric disorders
18108397	SPH 55.1 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy , Congenital Adrenal Hyperplasia,
18108398	SPH 55.2 (Fetal Lung)	Fetal Lung	Cystic Fibrosis, infection, lung cancer
18108364	SPH 55.3 (B's Lymphoma- Raji)	Lymphoma derived from B cells	
18108388	SPH 55.4 (Mammary Gland)	Mammary Gland	Lactation disorders, breast cancer
18108358	SPH 55.5 (Salivary Gland)	Salivary Gland	Dry mouth, infection
20281099	SPH 56.2 (MG63)		
20281100	SPH 56.3 (U5MC)		
264404	SPH.1 (Brain)	Whole Brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberculous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection

264510	SPH.10 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
264511	SPH.11 (Placenta)	Placenta	Infertility, birth defects
264512	SPH.12 (Thyroid)	Thyroid	Hyperparathyroidism, Hypoparathyroidism
264555	SPH.13 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264556	SPH.14 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264557	SPH.15 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264558	SPH.16 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264559	SPH.17 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264569	SPH.19 (One Fetal tissue and two cell lines)	Mixed	
264687	SPH.19.1 (fetal thymus - CRL7046)	Fetal Thymus	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, immunodeficiencies
264688	SPH.19.2 (hematopoietic stem cells - CRL2043)	Hematopoietic stem cells	Leukemia, osteoporosis, post-chemotherapeutic stem cell repopulation
264689	SPH.19.3 (osteogenic sarcoma cell lines - HTB96)	Osteogenic Sarcoma	Sarcomas, osteoporosis, osteopetrosis
264690	SPH.19.4 (Fetal Liver)	Fetal liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
264691	SPH.19.5 (Heart)	Heart	Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation
264692	SPH.19.6 (Spleen)	Spleen	Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host
264693	SPH.19.7 (Pituitary)	Pituitary	Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264482	SPH.2 (Brain)	Brain	Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection

264600	SPH.21 (Fetal Brain)	Fetal brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264601	SPH.22 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264602	SPH.23 (Thyroid)	Thyroid	Hyperthyroidism and Hypothyroidism
264603	SPH.24 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
264604	SPH.25 (Lymph Node)	Lymph Node	Lymphedema , Allergies
264605	SPH.26 (Placenta)	Placenta	Infertility, birth defects
264634	SPH.28 (Heart)	Heart	Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis , Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus , Pulmonary stenosis , Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation
264635	SPH.29 (Fetal Kidney)	Fetal Kidney	Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
264483	SPH.3 (Bone Marrow)	Bone marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264636	SPH.30 (Lymph Node)	Lymph Node	Lymphedema , Allergies
264637	SPH.31 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
264638	SPH.32 (Thyroid)	Thyroid	Hyperthyroidism and Hypothyroidism
264639	SPH.33 (Fetal Brain)	Fetal brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264484	SPH.4 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264758	SPH.44.1 (Kidney)	Kidney	Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
264760	SPH.44.2 (Fetal Liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation

264762	SPH.44.3 (Heart)	Heart	Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation
264764	SPH.44.4 (Prostate)	Prostate	Prostate Cancer
264766	SPH.44.5 (Spleen)	Spleen	Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host
264768	SPH.44.6 (pituitary)	Pituitary	Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264769	SPH.44.7 (Uterus)	Uterus	Infertility, birth defects
264905	SPH.48.1 (Burkitt's Lymphoma- Raji)	Burkitt's Lymphoma	Lymphoma, blood cancers
264906	SPH.48.2 (Thalamus- Brain)	Thalamus	Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264907	SPH.48.3 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy, Congenital Adrenal Hyperplasia,
264908	SPH.48.4 (Fetal Lung)	Fetal Lung	Cystic Fibrosis, infection, lung cancer
264909	SPH.48.5 (Salivary Gland)	Salivary Gland	Dry mouth, infection
264910	SPH.48.6 (Mammary Gland)	Mammary Gland	Lactation disorders, breast cancer
265006	SPH.50.1 (B's lymphoma)	Burkitt's Lymphoma	Lymphoma, blood cancers
265007	SPH.50.2 (thalamus)	Thalamus	Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
265008	SPH.50.3 (adrenal gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy, Congenital Adrenal Hyperplasia,
265009	SPH.50.4 (fetal lung)	Fetal Lung	Cystic Fibrosis, infection, lung cancer
265010	SPH.50.5 (salivary gland)	Salivary Gland	Dry mouth, infection
265011	SPH.50.6 (mammary gland)	Mammary Gland	Lactation disorders, breast cancer
18108385	SPH.51.1 (MCF-7)	Breast Cancer	Breast Cancer
18108370	SPH.51.2 (CCRF-CEM)	Cancer Cell line	Cancer
18108374	SPH.51.3 (K-562)	Cancer Cell line	Cancer
18108351	SPH.51.4 (OVCAR-3)	Ovarian cancer	Ovarian cancer
18108372	SPH.51.5 (HL-60)	Cancer Cell line	Cancer
264486	SPH.6 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,

264508	SPH.8 (Fetal Brain)		Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264509	SPH.9 (Lymph Node)	Lymph Node	Lymphedema , Allergies
20798451	SRH.56.3(UtSMC)		
264487	SRH.1 (Brain)	Brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264534	SRH.11 (Bone marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264535	SRH.12 (Bone marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264563	SRH.19 (Fetal Brain)	Fetal brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264488	SRH.2 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264564	SRH.20 (Lymph Node)	Lymph Node	Lymphedema , Allergies
264565	SRH.21 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
264566	SRH.22 (Placenta)	Placenta	Infertility, birth defects
264567	SRH.23 (Thyroid)	Thyroid	Hyperthyroidism and Hypothyroidism
264591	SRH.25 (Fetal Brain)	Fetal brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264592	SRH.26 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264593	SRH.27(thyroid)	Thyroid	Hyperthyroidism and Hypothyroidism
264594	SRH.28 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
264595	SRH.29 (Lymph Node)	Lymph Node	Lymphedema , Allergies
264489	SRH.3 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,

264596	SRH.30 (Placenta)	Placenta	Infertility, birth defects
264628	SRH.33 (fetal Kidney)	Fetal kidney	Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
264629	SRH.34 (lymph Node)	Lymph Node	Lymphedema , Allergies
264630	SRH.35 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
264631	SRH.36 (thyroid)	Thyroid	Hyperthyroidism and Hypothyroidism
264632	SRH.37 (Fetal Brain)	Fetal Brain	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264490	SRH.4 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
264681	SRH.43.1 (fetal thymus - CRL7046)	Fetal Thymus	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, immunodeficiencies
264682	SRH.43.2 (hematopoietic stem cells - CRL2043)	Hematopoietic stem cells	Leukemia, osteoporosis, post-chemotherapeutic stem cell repopulation
264683	SRH.43.3 (osteogenic sarcoma cell lines - HTB96)	Osteogenic Sarcoma	Sarcomas, osteoporosis, osteopetrosis
264684	SRH.43.4 (Fetal Liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
264685	SRH.43.6 (Spleen)	Spleen	Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura , Immunodeficiencies, Graft versus host
264686	SRH.43.7 (pituitary)	Pituitary	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264757	SRH.44.1 (Kidney)	Kidney	Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
264759	SRH.44.2 (Fetal Liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
264761	SRH.44.3 (Heart)	Heart	Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus , Pulmonary stenosis , Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation
264763	SRH.44.4 (Prostate)	Prostate	Prostate Cancer
264765	SRH.44.5 (Spleen)	Spleen	Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura , Immunodeficiencies, Graft versus host

264767	SRH.44.6 (Pituitary)	Pituitary	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
264828	SRH.46.1 (Lymph Node)	Lymph Node	Lymphedema , Allergies
264887	SRH.47.5 (Fetal Liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
18108377	SRH.50.1 (B's lymphoma)	Burkitt's Lymphoma	Lymphoma, blood cancers
18108380	SRH.50.2 (thalamus)	Thalamus	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
18108396	SRH.50.3 (adrenal gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy , Congenital Adrenal Hyperplasia,
18108391	SRH.50.4 (fetal lung)	Fetal Lung	Airway diseases, infection
18108357	SRH.50.5 (salivary gland)	Salivary Gland	Dry mouth, infection
18108390	SRH.50.6 (mammary gland)	Mammary Gland	Lactation disorders, breast cancer
264532	SRH.9 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
263974	736xN		
263976	736xN		
263981	736xN		
20281166	96xN		
20281169	96xN		
20281171	96xN		
263994	cDNA-ORF Selection		
264080	Mx96		
21906754	NQH 6.1 (HH729)		
22278996	NQH 6.10 (PrEC)	Endothelial cells	heart disease, cancer
22278997	NQH 6.11 (CAEC)	Endothelial cells	heart disease, cancer
22278998	NQH 6.12 (CSC)	Cancer Cell line	Cancer
22278999	NQH 6.13 (NHNPC)	Cancer Cell line	Cancer
22279000	NQH 6.14 (NHMC-RM)	Cancer Cell line	Cancer
22279002	NQH 6.15 (Hypothalamus)	Hypothalamus	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Obesity
21906764	NQH 6.2 (In Dated Platelets)	Platelets	Clotting diseases, stroke
21906765	NQH 6.3 (HuVec)	Endothelial cells	heart disease, cancer
87168474	NQH 6.3 (Sized-HUVEC)	Endothelial cells	heart disease, cancer
21906766	NQH 6.4 (U1MVEC- myo)	Cancer Cell line	Cancer
21906767	NQH 6.5 (NHEM-neo)	Cancer Cell line	Cancer
21906768	NQH 6.6 (NHEK)	Cancer Cell line	Cancer
21906769	NQH 6.7 (ByCAEC)	Endothelial cells	heart disease, cancer
22278994	NQH 6.8 (NHA)	Cancer Cell line	Cancer



22278995	NQH 6.9 (PrSC)	Cancer Cell line	Cancer
27486261	NQH 7.1 (Jurkat E6-untreated)	Cancer Cell line	Cancer
27486262	NQH 7.2 (TF1-untreated)	Cancer Cell line	Cancer
27486264	NQH 7.3 (U87-untreated)	Cancer Cell line	Cancer
27486265	NQH 7.4 (THP1-untreated)	Cancer Cell line	Cancer
29331822	NQH 8.1 (Brain- amygdala)		Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
29331824	NQH 8.2 (Brain-hippocampus)		Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
29331825	NQH 8.3 (Brain- substantia nigra)		Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection
29331826	NQH 8.4 (small intestine)	Small intestine	digestive diseases, obesity, diabetes
29331827	NQH 8.5 (Spinal cord)	Spinal chord	paralysis, neurodegenerative disorders
29331828	NQH 8.6 (stomach)	Stomach	Stomach cancer
29331830	NQH 8.7 (Trachea)	Trachea	Airway diseases, infection
87168518	NQH 9.1 (Sized-MG-63_treatment pool)		
87168559	NQH 9.2 (Sized-HEPG2 untreated)		
35695763	NQH.10.1 (MCF-7untreated)	Cancer Cell line	Cancer
35695855	NQH.10.2 (U-937_treatment pool)	Cancer Cell line	Cancer
35695917	NQH.10.3 (JAR)	Cancer Cell line	Cancer
35696052	NQH.10.4 (PA-1)	Cancer Cell line	Cancer
35696286	NQH.10.5 (CADMEC)	Endothelial cells	heart disease, cancer
35696423	NQH.10.6 (CADMEC_LA)	Endothelial cells	heart disease, cancer
52644045	NQH.11.1 (SK-PN-DW)	Cancer Cell line	Cancer
52644150	NQH.11.2 (Chorionic Villus Cells)	Chorionic villus	fertility, birth defects
52644229	NQH.11.3 (A549)	Cancer Cell line	Cancer
52644296	NQH.11.4 (U266B1)	Cancer Cell line	Cancer
52644332	NQH.11.5 (Daoy)	Cancer Cell line	Cancer
52644507	NQH.11.6 (SW1783)	Cancer Cell line	Cancer
52645080	NQH.12.1 (U-118MG)	Cancer Cell line	Cancer
52645129	NQH.12.2 (A204)	Cancer Cell line	Cancer
52645156	NQH.12.3 (T24)	Cancer Cell line	Cancer
52646317	NQH.12.4 (G-401)	Cancer Cell line	Cancer
52646365	NQH.12.5 (CaSki)	Cancer Cell line	Cancer
52646842	NQH.12.6 (SHP-77)	Cancer Cell line	Cancer

60424179	NQH.14.1 (Yale75_breast carcinoma)	Breast carcinoma	Breast Cancer
60424269	NQH.14.2 (Yale78B_ovarytumor)	Ovary tumor	Ovarian cancer
60431528	NQH.14.3 (Yale79_prostateBPH)	Prostate	Prostate Cancer
60431602	NQH.14.4 (Yale80_ProstateAdenocarcinoma)	Prostate	Prostate Cancer
60431735	NQH.14.5 (Yale86_UterineMyoma)	Uterine Myoma	Uterine Cancer
60431850	NQH.14.6 (Yale207_Myometrium)	Myometrium	Fertility
60432049	NQH.15.1 (Yale99_cervix)	Cervix	Osteoporosis, cervical cancer
60432113	NQH.15.2 (Yale45_spleenITP)		Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host
60432229	NQH.15.3 (Yale16_Skin)	Skin	wound healing, melanoma
60432289	NQH.15.4 (Yale137_Parotid)		
60433356	NQH.15.5 (Yale38_SmallIntestine)	Small intestine	digestive diseases, obesity, diabetes
60433438	NQH.15.6 (Yale28_ColonAscending)	Colon	Colon cancer
65274444	NQH.17.1 (Larynx)	Larynx	Cancer
65274572	NQH.17.2 (Duodenum)	Duodenum	
65274620	NQH.17.3 (Kidney, Primary tumors)		Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
65274727	NQH.17.4 (Lung Pleura, normal)	Lung	Airway diseases, infection
65274791	NQH.17.5 (Lung, Normal Adult)	Lung	Airway diseases, infection
83373044	NQH.18.230 (Pooled adrenal gland, placenta)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy, Congenital Adrenal Hyperplasia,
85658542	NQH.18.560 (Pooled uterus, BeWo pool)	Uterus	Infertility, birth defects
33656970	NQH.9.1 (MG-63_treatment pool)	Cancer Cell line	Cancer
33657023	NQH.9.2 (HEPG2 untreated)		Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
33657084	NQH.9.3 (PC3_untreated)	Cancer Cell line	Cancer
33657109	NQH.9.4 (TF-1_TPA)	Cancer Cell line	Cancer
33657182	NQH.9.5 (TF-1_TPO)	Cancer Cell line	Cancer
33657349	NQH.9.6 (TF-1_Hemin)	Cancer Cell line	Cancer
33657402	NQH.9.7 (HFDPC)	Cancer Cell line	Cancer
264259	NQH1 (Mixture of eight adult & two fetal tissues)		
264288	NQH2 (Ten tissues plus lymphocyte control)		
264448	NQH3 (Bone Marrow)	Bone Marrow	Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host,
265017	NQH4.1 (lymph node)	Lymph Node	Lymphedema, Allergies

265018	NQH4.2 (fetal kidney)	Fetal Kidney	Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
66712502	NQH4.2 (Sized)		
265019	NQH4.3 (pituitary gland)		Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Obesity
66714117	NQH4.3 (Sized)		
265020	NQH4.4 (testis)	testis	Infertility, birth defects
265021	NQH4.5 (fetal liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
265022	NQH4.6 (thyroid)	Thyroid	Hyperthyroidism and Hypothyroidism
18108376	NQH5.1 (MCF-7)	Breast cancer	Breast Cancer
18108387	NQH5.2 (CCR5-CEM)	Cancer Cell line	Cancer
264952	NRL1: HPLC FRACTIONATION OF RELIG		
263971	Old BB3 Baits		
263969	Old BB5 Baits		
263975	ORFSEL		
263972	OTHER Baits		
263978	pGALORF		
264106	PPBAITS		
264088	QC-YA7		
264089	QC-YA8		
264102	Resequenced Interactors		
264369	RRH.1		
60170394	RRH.10.1 (MCF-7untreated)	Breast cancer	Breast Cancer
60170615	RRH.10.2 (U-937_treatment pool)	Cancer Cell line	Cancer
60170831	RRH.10.3 (JAR)	Cancer Cell line	Cancer
60174639	RRH.11.8 (HeLa)	Cancer Cell line	Cancer
264113	rrQEA Baits		
263973	RRQEA_B5 baits		
29146498	SRD 3.1 (SKMC)	Cancer Cell line	Cancer
29146499	SRD 3.2 (SKMC)	Cancer Cell line	Cancer
29147620	SRD 3.3 (RPTEC)	Cancer Cell line	Cancer
29148627	SRD 3.4 (HRCE)	Cancer Cell line	Cancer
29148629	SRD 3.6 (HRE)	Cancer Cell line	Cancer
29148784	SRD 3.7 (HRE)	Cancer Cell line	Cancer
55810764	SRD.7.1 (Lymph Node)	Lymph Node	Lymphedema, Allergies
55811150	SRD.7.2 (pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
55811386	SRD.7.3 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy, Congenital Adrenal Hyperplasia,

55811576	SRD.7.4 (Pituitary Gland)	Pituitary	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Obesity
55811957	SRD.7.5 (Fetal Liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
55812038	SRD.7.6 (Fetal Kidney)	Fetal kidney	Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
56181562	SRD.8.1 (Lymph Node)	Lymph Node	Lymphedema , Allergies
56181686	SRD.8.2 (Pancreas)	Pancreas	Pancreatitis, diabetes, pancreatic cancer
56182181	SRD.8.3 (Adrenal Gland)	Adrenal Gland/Suprarenal gland	Adrenoleukodystrophy , Congenital Adrenal Hyperplasia,
56182323	SRD.8.4 (Pituitary Gland)	Pituitary	Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Obesity
56182435	SRD.8.5 (Fetal Liver)	Fetal Liver	Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation
56182575	SRD.8.6 (Fetal Kidney)		Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome
32833986	SRD4: HL adapter		
56526486	SRD5: lrr fragments		
33109954	SRD5: long-RXRJ		
56994075	SRD9.1 (CS/SC)	Cancer Cell line	Cancer
263977	TSC Screen I		

```

      370              375              380
Glu Gln Val Ile Asn Asn Ile Leu Glu Glu Arg Leu Ala Pro Thr Leu
385              390              395              400
Ser Gln Leu Asp Arg Asn Leu Asp Arg Glu Met Lys Pro Asp Pro Thr
      405              410              415
Pro Leu Leu Thr Ser Arg His Asn Val Phe Gln Asn Asp Glu Phe Asp
      420              425              430
Val Phe Ser Arg Asp Ser Val Asp Leu Ser Arg Val His Lys Gly Lys
      435              440              445
Ser Thr Arg Lys Glu Glu Asn Thr Arg Ser Leu Leu Asn Asp Lys Arg
      450              455              460
Ala Val Ala Ala Gln Arg Gln Arg Tyr Glu Gln Tyr Ser Val Val Val
465              470              475              480
Glu Glu Val Pro Leu Gln Pro Gly Glu Ser Leu Pro Tyr His Ser Val
      485              490              495
Tyr Tyr Glu Asp Glu Tyr Asp Asp Thr Tyr Asp Gly Asn Gln Val Gly
      500              505              510
Ala Asn Asp Ala Asp Ser Met Thr Ser Ser Ser Ala Ala Gly His Ser
      515              520              525
Pro Ser Gln Val Leu Arg Thr Lys Val Pro Arg Glu Gly Gln Glu Glu
      530              535              540
Asp Asp Asp Asp Glu Glu Asp Asp Ala Asp Glu Glu Ala Pro Lys Pro
545              550              555              560
Asp His Phe Val Gln Asp Pro Ala Val Leu Arg Glu Lys Ala Glu Ala
      565              570              575
Arg Arg Met Ala Phe Leu Ala Lys Lys Gly Tyr Arg His Asp Ser Ser
      580              585              590
Thr Ala Val Ala Gly Ser Pro Arg Gly His Gly Gln Ser Arg Glu Thr
      595              600              605
Thr Gln Glu Arg Arg Lys Lys Glu Ala Asn Lys Ala Thr Arg Ala Asn
      610              615              620
His Asn Arg Arg Thr Met Ala Asp Arg Lys Arg Ser Lys Gly Met Ile
625              630              635              640
Pro Ser

```

&lt;210&gt; 5413

&lt;211&gt; 1677

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5413

```

agagatgggt gtgtaatgaa aattacaagg tcgttgaaca aggttagtag tgtcttgcct
60
ttttattctg tcatctcaaa cttcataaac agtcatacgt tctttgaaac gtagatttaa
120
tgtgtgcagt catttataaa tcaatgacat ttctcttttt tgtcataaaa ctgtatactg
180
aagaaattaa cgaatgcaca gtttctaaag ctgttgcatt tgtctgtgga atcataggtt
240
cccactaaga agaatttcag cattctggcc agaaatttga atacaattca agttgaagaa
300
atgtctgcct gtaacattag catccagggt cccagcatat ataataagga gcctaaaaat
360

```

ataataaatc ctcatgaaaa agttcaaatg aagtcaattt gtgcaaattc tcctataaag  
 420  
 gcacaacagg atcaattaca agtaaaaaac aatataaaaag caagtcttca caatgtcaaa  
 480  
 agttccttac ctctttttta tactaagtcc tctacttctg tggggcagtt gcagtctcct  
 540  
 accttgaatt cacctatcta tatgcaaaaag caaggaaaaa atgaacatct tgcatttaat  
 600  
 accaaatcta aggcttcaac agttggttca gaattggtac ttgtttctac caccgttcca  
 660  
 actgttcac atgtttctga tttggaaatg agctctactc tggactgttt acctgtgttg  
 720  
 gctgattggg aggatgtggt tttactgcca gcatctcagc ctgaggaaaa cgtagactgt  
 780  
 acagttccca ttagtgactc agacttagag atttcattta attctggaga aagattaatg  
 840  
 gttttgaaag aattggaaat gtcaagtcac gaaaactttg gagacataga ggaaactcct  
 900  
 caaaaatctg agacttctaa gtctattgtg tacaagagtc ctcacactac tatttataat  
 960  
 gtaaaagaag ccaaagatcc aggttcagat atttctgcct ttaagttacc tgaacacaaa  
 1020  
 tcaagtacct tcaacagagt taatgccaat atgtctcacc ctttagtttt ggggaaacat  
 1080  
 cctcttcttt caggtggtac caaaaggaat ccatgcagtc cccaagcttt cccaccagca  
 1140  
 aaaaaacaac ccttactat tcatgaagaa aagcctacat catctgattg ctcccagta  
 1200  
 agaagttctt cctggaggcg tctcccatct atattaactt ctacagttaa cctacaagag  
 1260  
 ccatggaaga gtgggaaaaat gacacctcca ttatgcaagt gtggtcggag atctaagaga  
 1320  
 cttgttgttt ctaataatgg accgaacctt ggaaaagtct tctattgttg ccctatcggg  
 1380  
 aaataccaag aaaacagaaa atgttgtggt tatttcaaat gggaacaaac acttcaaaag  
 1440  
 gaaagagcca acagcatggt tccatctcat tccacagggg gactcacttt tagttctcca  
 1500  
 gaaacaagcc atatttgtga cagaaattta agtatttcca caaaaaatc tttgagactc  
 1560  
 aggccttcaa tgaggaattg ataaccttct atgtatgaat cctaattggt ccttgaattt  
 1620  
 ccaaacatga gtattctgat aacatcttac actattttat ttttatttta tatatta  
 1677

&lt;210&gt; 5414

&lt;211&gt; 426

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 5414

Met Ser Ala Cys Asn Ile Ser Ile Gln Gly Pro Ser Ile Tyr Asn Lys  
 1 5 10 15  
 Glu Pro Lys Asn Ile Ile Asn Pro His Glu Lys Val Gln Met Lys Ser

**What is claimed is:**

1. An isolated nucleic acid molecule encoding a polypeptide comprising an amino acid sequence that is at least 85% identical to a polypeptide including an amino acid sequence selected from the group consisting of SEQ ID NO:2 $n$ , wherein  $n$  is any integer 1-3161, or the complement thereof.
2. The isolated nucleic acid molecule of claim 1, said molecule hybridizing under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule comprising the sequence of nucleotides selected from the group consisting of SEQ ID NO:2 $n$ , wherein  $n$  is any integer 1-3161, or the complement thereof.
3. The isolated nucleic acid molecule of claim 1, said molecule encoding a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2 $n$ , wherein  $n$  is any integer 1-3161, or an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO: 2 $n$ .
4. The isolated nucleic acid molecule of claim 1, wherein said molecule encodes a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2 $n$ , wherein  $n$  is any integer 1-3161.
5. The isolated nucleic acid molecule of claim 1, wherein said molecule comprises the sequence of nucleotides selected from the group consisting of SEQ ID NO:2 $n$ -1, wherein  $n$  is any integer 1-3161, or the complement thereof.
6. An oligonucleotide less than 100 nucleotides in length and comprising at least 10 contiguous nucleotides selected from the group consisting of SEQ ID NO:2 $n$ -1, wherein  $n$  is any integer 1-3161, or the complement thereof.
7. A vector comprising the nucleic acid molecule of claim 1.

8. The vector of claim 7, wherein said vector is an expression vector.
9. A host cell comprising the isolated nucleic acid molecule of claim 1.
10. A substantially purified polypeptide comprising an amino acid sequence at least 80% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:  $2n$ , wherein  $n$  is any integer 1-3161.
11. The polypeptide of claim 10, wherein said polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NO:  $2n$ , wherein  $n$  is any integer 1-3161.
12. An antibody that selectively binds to the polypeptide of claim 10.
13. A pharmaceutical composition comprising a therapeutically or prophylactically effective amount of a therapeutic selected from the group consisting of:
  - a) the nucleic acid of claim 1;
  - b) the polypeptide of claim 10; and
  - c) the antibody of claim 12;and a pharmaceutically acceptable carrier.
14. A kit comprising in one or more containers, a therapeutically or prophylactically effective amount of the pharmaceutical composition of claim 13.
15. A method of producing the polypeptide of claim 10, said method comprising culturing the host cell of claim 9 under conditions in which the nucleic acid molecule is expressed.
16. A method of detecting the presence of the polypeptide of claim 10 in a sample, comprising contacting the sample with a compound that selectively binds to said polypeptide under conditions allowing the formation of a complex between said polypeptide and said



compound, and detecting said complex, if present, thereby identifying said polypeptide in said sample.

17. A method of detecting the presence of a nucleic acid molecule of claim 1 in a sample, the method comprising contacting the sample with a nucleic acid probe or primer that selectively binds to the nucleic acid molecule and determining whether the nucleic acid probe or primer bound to the nucleic acid molecule of claim 1 is present in the sample.

18. A method for modulating the activity of the polypeptide of claim 10, the method comprising contacting a cell sample comprising the polypeptide of claim 10 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

19. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a ORFX-associated disorder, wherein said therapeutic is selected from the group consisting of:

- a) the nucleic acid of claim 1;
- b) the polypeptide of claim 10; and
- c) the antibody of claim 12.

20. A method for screening for a modulator of activity or of latency or predisposition to an ORFX-associated disorder, said method comprising:

- a) contacting a test compound with the polypeptide of claim 10; and
- b) determining if said test compound binds to said polypeptide,

wherein binding of said test compound to said polypeptide indicates the test compound is a modulator of activity or of latency or predisposition to an ORFX-associated disorder.

21. A method for screening for a modulator of activity or of latency or predisposition to an ORFX-associated disorder, said method comprising:

- a) administering a test compound to a test subject at an increased risk ORFX-associated disorder, wherein said test subject recombinantly expresses a polypeptide encoded by the nucleotide of claim 1;

- b) measuring expression the activity of said protein in said test subject;
- c) measuring the activity of said protein in a control subject that recombinantly expresses said protein and is not at increased risk for an ORFX-associated disorder; and
- d) comparing expression of said protein in said test subject and said control subject, wherein a change in the activity of said protein in said test subject relative to said control subject indicates the test compound is a modulator or of latency of predisposition to an ORFX-associated disorder.

22. The method of claim 20, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

23. A method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide of claim 11 in a subject, the method comprising:

- a) measuring the amount of the polypeptide in a sample from said subject; and
- b) comparing the amount of said polypeptide in step (a) to the amount of the polypeptide present in a control sample,

wherein an alteration in the level of the polypeptide in step (a) as compared to the control sample indicates the presence of or predisposition to a disease in said subject.

24. The method of claim 23, wherein said subject is a human.

25. A method for determining the presence of or predisposition to a disease associated with altered levels the nucleic acid molecule of claim 1 in a subject, the method comprising:

- a) measuring the amount of the nucleic acid in a sample from the mammalian subject; and
- b) comparing the amount of said nucleic acid in step (a) to the amount of the nucleic acid present in a control sample,

wherein an alteration in the level of the nucleic acid in step (a) as compared to the control sample indicates the presence of or predisposition to said disease in said subject.

26. The method of claim 25, wherein said subject is a human.

27. A method of treating or preventing a pathological condition associated with an ORFX-associated disorder in a subject, the method comprising administering to said subject a polypeptide of claim 10 in an amount sufficient to alleviate or prevent said pathological condition.

28. The method of claim 27, wherein said subject is a human.

29. A method of treating or preventing a pathological condition associated with an ORFX-associated disorder in a subject, the method comprising administering to said subject a nucleic acid molecule of claim 1 in an amount sufficient to alleviate or prevent said pathological condition.

30. The method of claim 29, wherein said subject is a human.

31. A method of treating or preventing a pathological condition associated with an ORFX-associated disorder in a subject, the method comprising administering to said subject an antibody of claim 12 in an amount sufficient to alleviate or prevent said pathological condition.

32. The method of claim 31, wherein said subject is a human.